Economics Bulletin

Volume 33, Issue 4

Testing the null of cointegration with a structural break: optimal kernel and bandwidth selection

Alexander Ludwig
TU Dresden

Abstract

Structural breaks in relationships between macroeconomic and financial time series are likely a result of financial crises or local reforms. If such structural breaks exist, cointegration tests have to take them into account. Arai and Kurozumi (2007), Carrion-i-Silvestre and Sanso (2006) and Kejriwal (2008) propose a test for the null of cointegration with structural breaks against the alternative of no cointegration (ACK test). In this paper, we systematically examine the ACK test along several dimensions: sample size, kernel selection, bandwidth selection, the maximum value of the first order autocorrelation coefficient allowed in the automated bandwidth estimators proposed by Andrews (1991) and break height. We then compare statistical error frequencies to those of the test proposed by Gregory and Hansen (1996) for the null of no cointegration against the alternative of cointegration with a structural break. We find that the ACK test performs better than the Gregory and Hansen (1996) test in some cases, especially when the data generating process does not contain a cointegration vector. If there is a cointegration vector, the ACK test usually leads to larger statistical error frequencies. The ACK test should hence be used in combination with other tests for cointegration.

I am grateful to Hans-Martin Ihle for helpful comments. Financial support by the PhD program of zeb/rolfes.schierenbeck.associates is gratefully acknowledged. The usual disclaimers apply.
Citation: Alexander Ludwig, (2013) ''Testing the null of cointegration with a structural break: optimal kernel and bandwidth selection'', *Economics Bulletin*, Vol. 33 No. 4 pp. 2828-2839.
Contact: Alexander Ludwig - alexander.ludwig@mailbox.tu-dresden.de.
Submitted: February 22, 2013. **Published:** December 03, 2013.

1. Introduction

The study of long-run equilibrium (cointegration) relationships between time series is one of the core empirical devices in macroeconomic and financial research. Testing for long-run equilibria requires long time series data, which is in turn likely to include structural breaks induced by economic crises and local political or economic reforms. If breaks are present but not considered in the model, tests for cointegration have low power (Perron, 2006).

As a complement to Gregory and Hansen's (1996) test (*GH test*) for the null of no cointegration against the alternative of cointegration with a structural break, Arai and Kurozumi (2007), Carrion-i-Silvestre and Sanso (2006) and Kejriwal (2008) propose the same test statistic for the null of cointegration with a structural break (*ACK test*).[1] Arai and Kurozumi (2007, p. 706) state that "from the view of classical hypothesis testing, if we are primarily concerned about cointegration with a structural break, [this] seems a more natural choice for the null hypothesis". Their test is an LM test which requires the estimation of a heteroscedastic and autocorrelation consistent error variance. To estimate this variance, Arai and Kurozumi (2007) and Carrion-i-Silvestre and Sanso (2006) apply the Bartlett kernel, whereas Kejriwal (2008, p. 14) does not further specify the "consistent estimate of the long-run variance".

Arai and Kurozumi (2007) analyze the size and power of the ACK test for two sets of data generating processes (*DGP*s) where the second set is one of the standard DGPs used in the cointegration literature since it emerged in the 1980s (by e.g. Banerjee et al., 1986, Engle and Granger, 1987, Gregory and Hansen, 1996). Arai and Kurozumi (2007, p. 725) conclude that the ACK test with Bartlett kernel "does not work well" on this set of DGPs in terms of size and power.[2]

In this paper, we use the same standard set of DGPs and analyze the size and power of the ACK test when using different kernels and bandwidth estimators to estimate the variance. We arrive at a more differentiated pattern of results. We first show that Andrews' (1991) automated data-dependent bandwidth estimator is hardly compatible with the ACK test, neither in its original version nor in the adjusted version in which a maximum value for the empirical first order autocorrelation coefficient of residuals is set. To address this issue, bandwidth estimators that only depend on the sample size (cf. Schwert, 1989) can be used.

We then compare the statistical error frequencies of the ACK and GH tests. The ACK test appears to be more reliable in some cases, especially when the true DGP does not contain cointegrated time series. Yet, in many cases the GH test outperforms the ACK test. The GH test suffers from power against the presence of structural breaks in regressions of non-cointegrated time series. We show that the application of the ACK test in turn implies a similar problem as size distortions increase with rising break heights.

The paper is structured as follows. In Section 2, we describe the ACK test and kernels used to estimate the long-run error variance. We then explain the model setup for analyzing the size and power of the test. Results for the case of data-dependent automated bandwidth

[1] According to Arai and Kurozumi (2007), the test was independently developed by these authors and Carrion-i-Silvestre and Sanso (2006). Arai and Kurozumi (2007, p. 707) point out that as a further contribution their "theory covers both a large and small structural change" and that they also give "assumptions in details and rigorous proof of the theorem for the unknown break point case ... for the dynamic least squares approach". Kejriwal (2008) generalized this test to a test of cointegration with multiple breaks. Models of two or more breaks are not within the scope of this paper.

[2] Carrion-i-Silvestre and Sanso (2006) use a similar DGP to assess the performance of the test statistic.

estimators are provided in Section 3 and for sample size-dependent bandwidths in Section 4. This section also includes the comparative assessment with respect to the GH test. Section 5 summarizes our conclusions.

2. Description of the ACK test and model setup

The ACK test is a test for the null of cointegration with a structural break at time T_1 against the alternative of no cointegration. The test statistic is based on the residuals of the regression

$$y_t = \mu_i + \beta_i' x_t + \sum_{j=-J}^{J} \pi_j \Delta x_{t-j} + e_t \quad \text{if } T_{i-1} < t \leq T_i \tag{1}$$

for $i = 1, 2,\ T_0 = 0,\ T_2 = T$ where J leads and lags of the regressor are included to correct for potential endogeneity (dynamic OLS estimation, DOLS). The number of leads and lags of the regressor can be determined in a specific-to-generic approach as in Arai and Kurozumi (2007). Using this approach, J is first set to $l_4 = \lfloor 4 * (T/100)^{1/4} \rfloor$ and then reduced sequentially until an F-test shows significance of the lead and lag with the highest order at the 10% level. The test statistic is calculated as

$$V = \frac{1}{T^2} \sum\nolimits_{i=1}^{T} \tilde{S}_i^2 / \tilde{\sigma}_e^2 \tag{2}$$

where $\tilde{S}_i = \tilde{e}_1 + \cdots + \tilde{e}_i$ is a sequence of partial sums of errors and $\tilde{\sigma}_e^2$ is a heteroscedastic and autocorrelation consistent estimator of the long-run variance:

$$\tilde{\sigma}_e^2 = \sum\nolimits_{j=-h}^{h} w(j,h) \frac{1}{T} \sum\nolimits_{t=1}^{T-j} \tilde{e}_t \tilde{e}_{t+j} \tag{3}$$

with kernel function $w(i, h)$ and bandwidth h. In this paper we apply a set of commonly used kernel functions (see Table 1). To estimate the bandwidth we use

1) automated bandwidth estimators derived by Andrews (1991) (see Table 1) and
2) sample size-dependent bandwidth estimators proposed by Schwert (1989) and Kwiatkowski et al. (1992): l_4, l_8 and l_{12} where $l_m = \lfloor m * (T/100)^{1/4} \rfloor$

Table 1: Kernel functions

Kernel	Kernel function $w(i,h) = w(i/h)$	Automated bandwidth estimator $\hat{h}$ (Andrews, 1991)
Barlett	$= \begin{cases} 1 - \lvert x \rvert & \text{for } \lvert x \rvert \leq 1 \\ 0 & \text{otherwise} \end{cases}$	$1.1447 * (\hat{a}(1) * T)^{1/3}$
Uniform	$= \begin{cases} 1 & \text{for } \lvert x \rvert \leq 1 \\ 0 & \text{otherwise} \end{cases}$	$0.6611 * (\hat{a}(2) * T)^{1/5}$
Tukey-Hanning	$= \begin{cases} (1 + \cos(\pi x))/2 & \text{for } \lvert x \rvert \leq 1 \\ 0 & \text{otherwise} \end{cases}$	$1.7462 * (\hat{a}(2) * T)^{1/5}$
Parzen	$= \begin{cases} 1 - 6x^2 + 6\lvert x \rvert^3 & \text{for } 0 \leq \lvert x \rvert \leq 0.5 \\ 2(1 - \lvert x \rvert)^3 & \text{for } 0.5 \leq \lvert x \rvert \leq 1 \\ 0 & \text{otherwise} \end{cases}$	$2.6614 * (\hat{a}(2) * T)^{1/5}$
Quadratic Spectral (QS)	$= \frac{25}{12\pi^2 x^2} \left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right)$	$1.3221 * (\hat{a}(2) * T)^{1/5}$

Note: $\hat{a}(1) = 4\hat{\rho}_A^2/((1+\hat{\rho}_A)^2(1-\hat{\rho}_A)^2)$ and $\hat{a}(2) = 4\hat{\rho}_A^2/(1-\hat{\rho}_A)^4$ where $\hat{\rho}_A$ is the first order autocorrelation coefficient of the residuals $\tilde{e}_t$ of eq. (1).

Critical values of the ACK test depend on the number of regressors and the location of the break. They are tabulated for break fractions that are multiples of 0.1 by Arai and Kurozumi (2007, p. 717) and Carrion-i-Silvestre and Sanso (2006, p. 629-30). For other break fractions, we simulate critical values via a discretization of the Wiener processes involved in the asymptotic distribution of the test (cf. Arai and Kurozumi, 2007, p. 716) by partial sums of i.i.d. normal random variables and 10.000 replications.

To analyze the size and power of the ACK test, we use the following data generating processes which are standard in the econometrics literature on cointegration (see e.g. Banerjee et al., 1986, Granger and Engle, 1987, Gregory and Hansen, 1996):

$$\begin{aligned} y_t + x_t &= v_t \qquad v_t(1-L) = \epsilon_{1t} \\ y_t + (2+\beta_t)x_t + 1 &= u_t \qquad u_t(1-\rho L) = \epsilon_{2t} \end{aligned} \tag{4}$$

where ϵ_{it} are independently $N(0,1)$-distributed random variables and L denotes the lag operator. If $\rho < 1$, there is a linear combination of x_t and y_t such that the two series are cointegrated, i.e. residuals u_t are stationary. If $\rho = 1$, x_t and y_t are not cointegrated. For this DGP, Arai and Kurozumi (2007, p. 725) report rejection frequencies larger than 0.2 for the case $\rho = 0$ (theoretical size: 0.05) and for some bandwidths non-monotonous rejection frequencies with respect to ρ and hence conclude that their test "does not work well" for this set of DGPs. The results provided in this paper will show a more differentiated pattern.

Unless stated otherwise, we set $\beta_t = 0$ for $t = 1, \cdots, T/2$ and $\beta_t = 2$ for $t = T/2+1, \cdots, T$. All experiments for the case of an unknown break date are replicated 1,000 times and for the case of a known break date 10,000 times. For the former, in each replication we estimate the break date T_1 by minimizing the sum of squared residuals:

$$\hat{T}_1 = \arg\min_{T_1} \sum_{i=1}^{2} \sum_{t=T_{i-1}}^{T_i} \left[y_t - \hat{a}_i - \hat{b}_i x_t \right]^2 \tag{5}$$

where $T_0 = 0$, $T_2 = T$ and $\hat{a}_i$ and $\hat{b}_i$ are least squares estimates in the respective regimes. Kejriwal and Perron (2008) show that this estimator of the break date is consistent even when x_t and y_t are non-stationary. We then compute the test statistic as described above and compare its value with the critical value at the 5% level (nominal size) corresponding to the break fraction $\hat{\lambda} = \hat{T}_1/T$.

3. Properties of the ACK test using Andrews' (1991) bandwidth estimators

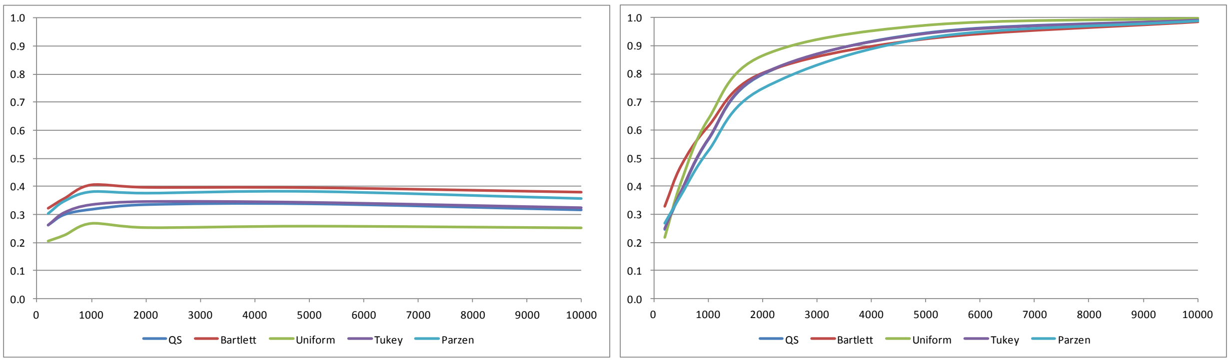
We first address the properties of the ACK test when the automated bandwidth estimators derived by Andrews (1991) (see Table 1) are used.

Consistency of the ACK test

Arai and Kurozumi (2007) propose to restrict the parameter $\hat{\rho}_A$ in Andrews' (1991) automated bandwidth estimators to $\rho_A^{max} = 0.9$ if $\hat{\rho}_A > 0.9$. In a first experiment, we illustrate the relevance of this adjustment. Figure 1 shows the power of the ACK test for both the non-adjusted and adjusted case with the five kernel functions considered and sample sizes between 200 and 10,000. In the non-adjusted case, rejection frequencies do not converge to 100%, i.e. the test is not consistent, whereas in the adjusted case they do converge. This is explainable by the fact that the series of residuals contains a unit root under the alternative. In this case, the bandwidth parameter h is of order T and the test statistic of order T/h (see Kurozumi, 2002). Hence, the test statistic does not diverge under the alternative and is therefore not consistent if the automated bandwidth parameter is used without any restrictions.

In the case of adjusted bandwidths, the ACK test is consistent. The speed of convergence, however, is relatively slow. More than 4,000 observations are required to achieve a power of more than 0.9 (see the right panel of Figure 1).

Figure 1: Power of the ACK test with automated bandwidth estimators w.r.t. sample size



Note: in the left panel Andrews' (1991) automated bandwidth estimates are used without adjustment (see Table 1), in the right panel the first order autocorrelation coefficient $\hat{\rho}_A$ is restricted to 0.9

The test statistic (2) hence depends on the additional parameter ρ_A^{max}. We will take this into account in the following examination of the size and power of the ACK test.

Size and power of the ACK test

We will now assess the size and power of the ACK test by varying the cap ρ_A^{max} for the maximum first order autocorrelation coefficient $\hat{\rho}_A$ in Andrews' (1991) automated bandwidth estimator. Here and in the following sections, we fix the sample size at $T = 200$. Tables 2 and 3 contain our results for the rejection frequencies of the ACK test under the null hypothesis, for which we choose $\rho = 0$ and $\rho = 0.5$ (as in Arai and Kurozumi, 2007), and under the alternative, i.e. for $\rho = 1$. The following conclusions can be drawn:

1) In all cases there is an over-rejection of the null. Size distortions increase with the AR parameter ρ in the DGP (4). This observation is due to the large parameter set attainable for ρ under the null. Rejection frequencies increase with ρ and converge to the rejection frequency under the alternative (see also Figure 2 below).

2) For $\rho_A^{max} = 0.9$ as proposed by Arai and Kurozumi (2007), the power of the test is relatively low. For $\rho_A^{max} = 0.8$ as proposed by Carrion-i-Silvestre and Sanso (2006), the power is not considerably larger.

3) The power of the test increases with a shrinking value of the maximum first order autocorrelation coefficient ρ_A^{max}. In the case of positively autocorrelated time series, a smaller bandwidth leads to smaller estimates of the long-run variance (3) and hence to larger values of the test statistic (2) and to larger rejection frequencies.

4) In comparison with the case of a known break location, the power of the test is smaller when the break location is unknown (cf. a corresponding finding by Arai and Kurozumi, 2007, p. 725). Here, the error incurred when estimating the location of the break has a further impact on the power of the test. Although Arai and Kurozumi (2007) show that the asymptotic distribution of the test statistic (2) under the null is the same in

both the known and the unknown break case, small sample sizes lead to substantial differences in the power of the test.[3]

5) The size of the test is similar for both the known and the unknown break case.

6) The choice of the kernel does not alter the results considerably. This is due to the kernel-individual optimization of the bandwidth estimator with respect to a mean squared error criterion (see Andrews, 1991).

Table 2: Size and power of the ACK test (automated bandwidth estimators, known break case)

ρ_A^{max} / ρ	1.0	0.9	0.8	0.7	0.6	0.5	0.4
Quadratic Spectral kernel							
0.0	0.098	0.106	0.102	0.109	0.106	0.103	0.104
0.5	0.132	0.136	0.134	0.136	0.140	0.147	0.159
1.0	0.394	0.396	0.437	0.535	0.622	0.707	0.779
Bartlett kernel							
0.0	0.130	0.126	0.118	0.121	0.128	0.127	0.127
0.5	0.171	0.172	0.173	0.170	0.169	0.190	0.206
1.0	0.459	0.466	0.525	0.595	0.665	0.739	0.820
Uniform kernel							
0.0	0.097	0.092	0.097	0.098	0.097	0.100	0.095
0.5	0.120	0.122	0.118	0.115	0.120	0.137	0.195
1.0	0.265	0.302	0.453	0.607	0.687	0.782	0.885
Tukey-Hanning kernel							
0.0	0.116	0.114	0.108	0.113	0.108	0.116	0.107
0.5	0.141	0.139	0.144	0.139	0.131	0.162	0.200
1.0	0.397	0.356	0.445	0.555	0.644	0.751	0.850
Parzen kernel							
0.0	0.107	0.114	0.103	0.112	0.108	0.114	0.108
0.5	0.131	0.133	0.138	0.134	0.141	0.147	0.175
1.0	0.439	0.399	0.429	0.509	0.611	0.682	0.784

Note: $T = 200$

Selection of the maximum first order autocorrelation coefficient

As seen above, reducing the value of the maximum first order autocorrelation coefficient ρ_A^{max} increases the power of the test. Looking at the results in Table 2, one may argue that the selection of $\rho_A^{max} = 0.7$ may be a good choice as a compromise between gains in power and avoidance of further size distortions. However, if the true autocorrelation coefficient ρ in the DGP (4) is larger than 0.7, size distortions will increase considerably. In Figure 2, we illustrate rejection frequencies depending on the true value of ρ in the DGP (4) and for each selection of the cap ρ_A^{max} in the test statistic.[4] For each value of ρ_A^{max}, size distortions are large for $\rho_A^{max} \leq \rho < 1$. In order to avoid additional size distortions, $\rho_A^{max} > \rho$ should hold. Yet, as the true autocorrelation coefficient ρ is unknown, the selection of ρ_A^{max} poses

[3]In the case that the break point estimator (5) yields a break location different from the middle of the sample, the corresponding critical value of the test becomes larger (cf. Arai and Kurozumi (2007), Table 1) which implies a reduction of the rejection frequency of the test, cet. par.

[4]As the results of the unknown and known break case are qualitatively the same, we only present the results for the known break case here. Results for other kernels than the quadratic spectral kernel do not differ considerably and can be obtained upon request.

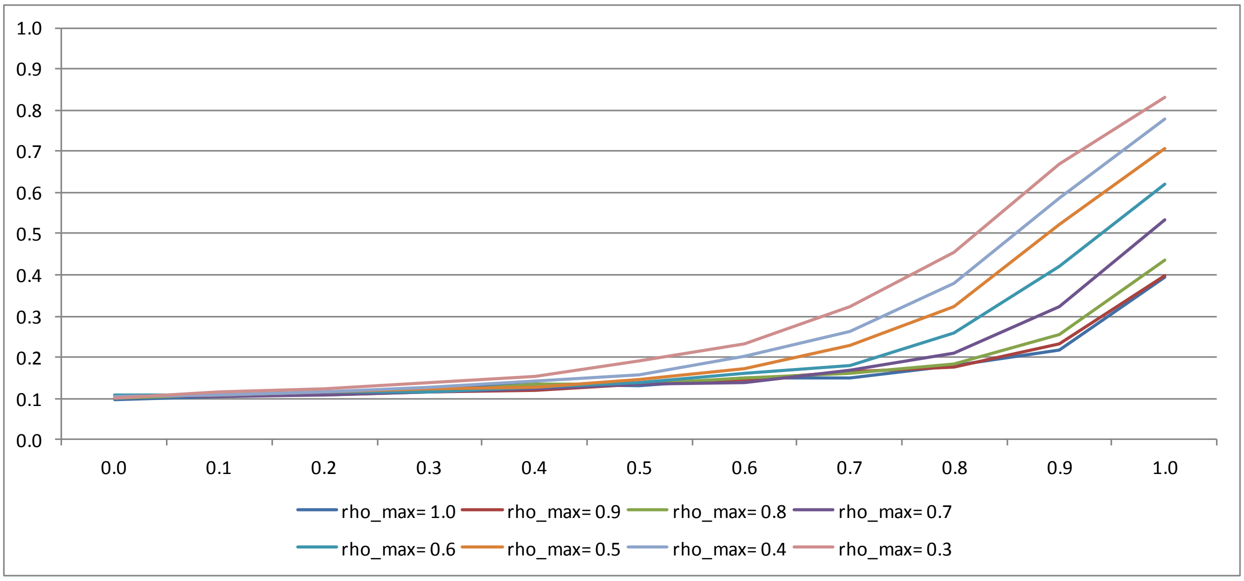
a major problem for making the test practical in combination with the data-dependent automated bandwidth estimators proposed by Andrews (1991).[5]

Table 3: Size and power of the ACK test (automated bandwidth estimators, unknown break case)

ρ_A^{max} / ρ	1.0	0.9	0.8	0.7	0.6	0.5	0.4
Quadratic Spectral kernel							
0.0	0.112	0.113	0.106	0.107	0.094	0.082	0.106
0.5	0.131	0.121	0.116	0.100	0.136	0.128	0.175
1.0	0.257	0.275	0.300	0.389	0.443	0.567	0.609
Bartlett kernel							
0.0	0.130	0.125	0.125	0.125	0.108	0.101	0.115
0.5	0.171	0.150	0.144	0.128	0.166	0.154	0.215
1.0	0.324	0.331	0.363	0.458	0.492	0.586	0.660
Uniform kernel							
0.0	0.107	0.099	0.105	0.105	0.088	0.084	0.096
0.5	0.121	0.099	0.104	0.090	0.115	0.124	0.199
1.0	0.201	0.216	0.318	0.461	0.509	0.645	0.747
Tukey-Hanning kernel							
0.0	0.117	0.117	0.112	0.111	0.096	0.085	0.109
0.5	0.134	0.123	0.116	0.105	0.137	0.140	0.203
1.0	0.262	0.256	0.304	0.405	0.460	0.609	0.702
Parzen kernel							
0.0	0.115	0.116	0.111	0.106	0.095	0.086	0.108
0.5	0.139	0.125	0.118	0.108	0.133	0.130	0.183
1.0	0.299	0.278	0.291	0.381	0.424	0.554	0.623

Note: $T = 200$

Figure 2: Rejection frequencies w.r.t. ρ in DGP (4)



Note: $T = 200$, QS kernel

[5]This selection problem also arises in the univariate version of this test, the unit-root test proposed by Kurozumi (2002). For this unit root test, size distortions also increase when the bandwidth-relevant autocorrelation parameter ρ_A^{max} is set below the true autocorrelation parameter ρ (see Kurozumi, 2002, Table 3).

4. Results of the ACK test using sample size-dependent bandwidth estimators

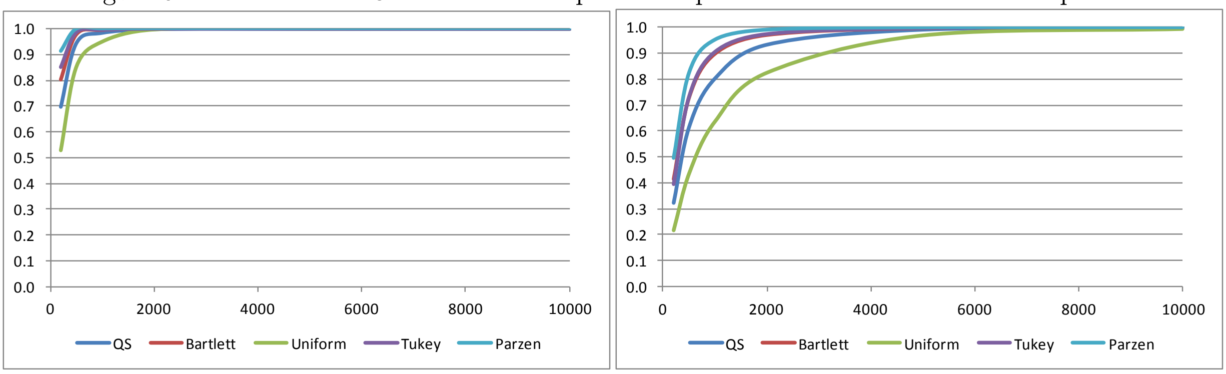
We now consider the sample size-dependent bandwidth estimators proposed by Schwert (1989) and Kwiatkowski et al. (1992) denoted by l_4, l_8 and l_{12} where $l_m = \lfloor m*(T/100)^{1/4} \rfloor$.

Consistency of the ACK test

For all three bandwidth estimators, l_4, l_8 and l_{12}, the ACK test is consistent. Figure 3 shows that the power of the test for bandwidths l_4 and l_{12} converges to 1 relatively fast. For example, using bandwidth l_4, the rejection frequencies under the alternative are larger than 0.95 for sample sizes of 500 (except for the uniform kernel). The graphs for bandwidth l_8 lie in between those of l_4 and l_{12} (not reported).

The power of the test is considerably higher when sample size-dependent bandwidth estimators are used to estimate the variance which is due to smaller bandwidths. An increase in the autocorrelation coefficient ρ in the DGP (4) does not influence the selection of the bandwidths in this case, whereas it leads to a rise in the data-dependent bandwidth estimators proposed by Andrews (1991).

Figure 3: Power of the ACK test with sample size-dependent bandwidths w.r.t. sample size



Note: in the left (right) panel the bandwidth estimator l_4 (l_{12}) is used

Size and power of the ACK test

In the following, we again fix the sample size at $T = 200$. Table 4 contains rejection frequencies of the test depending on the true autocorrelation coefficient ρ in DGP (4), the kernel and the sample size-dependent bandwidth. The following conclusions can be drawn:

1) Results vary considerably with respect to the kernel. This results from the kernel-independent selection of the bandwidth which works well only for some of the kernels. This is in contrast to the data-dependent bandwidth estimator proposed by Andrews (1991) which is optimized for each kernel separately (cf. Table 1).

2) Rejection frequencies increase with ρ for all kernels and bandwidths except for the uniform kernel applied with bandwidths l_8 and l_{12}. This is a similarly unexpected result as those shown by Arai and Kurozumi (2007, p. 725) for this DGP.

3) Increasing the bandwidth reduces the power of the test. This results from the same argument as in conclusion 3) in Section 3.

4) Compared with the results for the data-dependent bandwidth selection (cf. Table 3), size distortions are similarly large for the quadratic spectral kernel and the uniform kernel here. For the other kernels size distortions are larger.

5) For all bandwidths, using the quadratic spectral kernel rather than the uniform kernel increases the power of the test.

In the following, we will focus on the quadratic spectral kernel. A conservative test for cointegration would use the smallest bandwidth l_4 because the error of not rejecting the null of cointegration although it should be rejected is smaller than for bandwidths l_8 and l_{12}.

Table 4: Size and power of the ACK test (sample size-dependent bandwidths)

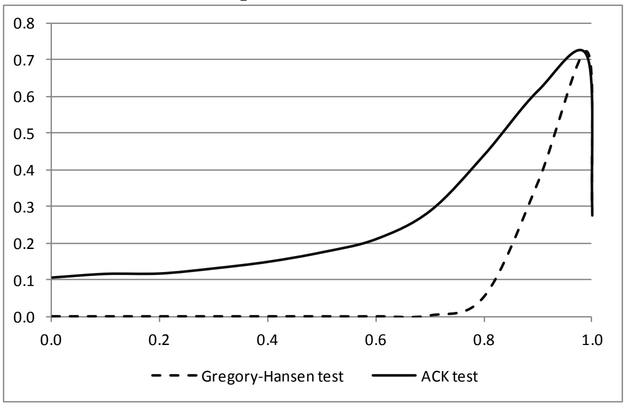
Bandwidth h / ρ	l_4	l_8	l_{12}
Quadratic Spectral kernel			
0.0	0.106	0.107	0.140
0.5	0.176	0.124	0.151
1.0	0.713	0.420	0.322
Bartlett kernel			
0.0	0.132	0.118	0.121
0.5	0.248	0.167	0.164
1.0	0.801	0.541	0.413
Uniform kernel			
0.0	0.103	0.179	0.308
0.5	0.111	0.140	0.280
1.0	0.527	0.265	0.216
Tukey-Hanning kernel			
0.0	0.132	0.100	0.108
0.5	0.277	0.143	0.137
1.0	0.849	0.556	0.395
Parzen kernel			
0.0	0.164	0.107	0.098
0.5	0.386	0.184	0.148
1.0	0.915	0.679	0.496

Comparison with size and power of the Gregory-Hansen (1996) test

We now compare the performance of the ACK test with the cointegration test proposed by Gregory and Hansen (1996) which is often used when structural breaks are taken into account in testing for cointegration of two time series. This test is a test for the null of no cointegration against the alternative of cointegration with a structural break. Gregory and Hansen suggest three different statistics but recommend the Z_t statistic which "appears best in terms of size and power" (cf. Gregory and Hansen, 1996, p. 114). We only use the Z_t statistic in this paper. Critical values do not depend on the location of the break and are tabulated in the original paper.

Figure 4 presents the statistical error frequencies for both tests. For autocorrelation parameters $\rho < 0.7$, the power of the GH test is 1, yielding an error of zero, whereas the size and hence the error of the ACK test exceeds 0.1. The error frequency graph is considerably steeper for the GH test. For autocorrelation parameters close to 1, errors are similar for both tests. The size of the GH test is 0.276 in this case which is due to the fact that this test also has power against the presence of structural breaks in regressions of non-cointegrated time series (see below). In sum, the GH test seems to outperform the ACK test over the interval $0 \leq \rho < 1$, i.e. when the true DGP contains cointegrated time series.

Figure 4: Statistical error frequencies of the GH and ACK tests w.r.t. ρ



Note: the graph displays the frequencies of a rejection of the null in the case that it is correct and of the acceptance of the null in the case it should be rejected.

Monotonicity of size and power with respect to break heights

Various works discuss the undesired property of non-monotonous power or accelerating size distortions with respect to increasing break heights (see Juhl and Xiao, 2009, for a univariate example and Kejriwal and Perron, 2010, for a multivariate case). It is well-known that the GH test also has power against the existence of structural breaks in regressions of time series that are not cointegrated. To visualize this, we set $\rho = 1$, which is the only value for ρ under the null of the GH test. Further, we model increasing break heights β_t for the second half of the sample ranging from $\beta_t = 0$ (no break) to $\beta_t = 8$. The results in the first row of Table 5 exhibit increasing size distortions for increasing break heights.

Further, we determine the power of the GH test given an autocorrelation parameter of $\rho = 0.9$ in DGP (4) and the same break heights as before. The results in the second row of Table 5 indicate a monotonous relation of the power of the test with respect to the break height. As we have argued before, the pure existence of a break increases the probability of rejecting the null. This monotonous relationship is thus not unexpected.

Returning to the ACK test, we now report rejection frequencies for the same autocorrelation parameters. The results in the third and fourth row of Table 5 show a positive relation of the power and the size of the ACK with increasing break heights. As the break height increases, the error in estimating the break location decreases. As discussed in conclusion 4) in Section 3, rejection frequencies are smaller in the presence of large uncertainty about the break location. Size distortions incurred by the application of the ACK test hence increase with the break height.

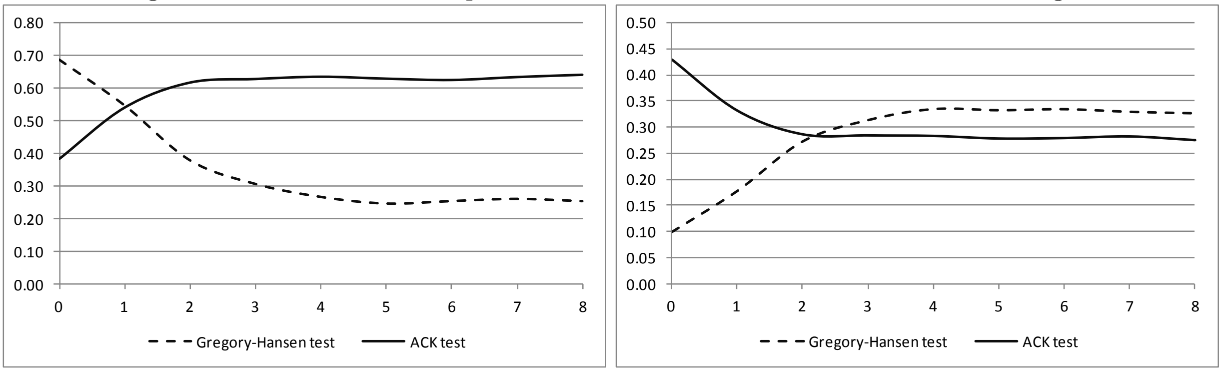
We now compare statistical error frequencies incurred by the application of both tests. In the case of cointegrated time series, the ACK test only performs better if there is either no break at all or only a small break in the relationship between both series (see the left panel in Figure 5). However, testing for the presence of a structural break in the regression of possible non-stationary series should be a prerequisite before using any of the cointegration tests discussed here. For breaks larger than $\beta_t = 1$, the GH test leads to considerably smaller statistical errors.

For non-cointegrated time series, the GH test performs better in the presence of small breaks (see the right panel in Figure 5). If break heights rise above $\beta_t = 2$, the ACK test leads to slightly lower error frequencies.

Table 5: Size and power of the GH and ACK test w.r.t. break heights

β_t	0.0	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0
Gregory-Hansen Z_t test									
$\rho = 1.0$ (size)	0.099	0.177	0.276	0.309	0.333	0.331	0.319	0.308	0.313
$\rho = 0.9$ (power)	0.311	0.452	0.632	0.694	0.734	0.730	0.736	0.740	0.747
ACK test (QS kernel, bandwidth l_4)									
$\rho = 1.0$ (power)	0.571	0.668	0.713	0.715	0.703	0.709	0.720	0.717	0.724
$\rho = 0.9$ (size)	0.383	0.541	0.618	0.629	0.636	0.630	0.626	0.635	0.642

Figure 5: Statistical error frequencies of the GH and ACK tests w.r.t. break heights



Note: the left and right panel consider the cases $\rho = 0.9$ (cointegration) and $\rho = 1$ (no cointegration), respectively.

5. Conclusions

Having tested for the existence of structural breaks in the relationship of non-stationary time series, it seems a straightforward next step to apply a test for the null of cointegration with a structural break against the alternative of no cointegration. Such a test was developed by Arai and Kurozumi (2007), Carrion-i-Silvestre and Sanso (2006) and Kejriwal (2008). It complements the widely used test proposed by Gregory Hansen (1996) where the null and alternative of the test are set in the reversed order.

The properties of the ACK test depend on the choice of the kernel and associated bandwidth estimators for the long-run variance. In this paper, we show that the automated bandwidth estimators proposed by Andrews (1991) are hardly compatible with this test. Sample size-dependent bandwidths as suggested by Schwert (1989) and Kwiatkowski et al. (1992) may be a better choice.

We showed that the GH test outperforms the ACK test in many cases. As the ACK test, however, involves a larger degree of conservatism with respect to inferring the presence of cointegration, especially when using the quadratic spectral kernel with Schwert's (1989) bandwidth l_4, we recommend to use this test in combination with the GH test. Alternatively, it could be complemented with standard tests for cointegration that are separately applied in each regime before and after the break. This can also help to counter the relatively large impact breaks have on the statistical errors incurred by the GH test when the true data generating process does not involve cointegration.

References:

Andrews, D.W.K. (1991) "Heteroskedasticity and autocorrelation consistent covariance matrix estimation" *Econometrica* **59**, 817-858.

Arai, Y. and E. Kurozumi (2007) "Testing for the null hypothesis of cointegration with a structural break" *Econometric Reviews* **26**, 705-739.

Banerjee, A., Dolado, J.J., Hendry, D.F. and G.W. Smith (1986) "Exploring equilibrium relationships in econometrics through static models: Some Monte Carlo evidence" *Oxford Bulletin of Economics and Statistics* **48**, 253-277.

Carrion-i-Silvestre, J.L. and A. Sanso (2006) "Testing the null of cointegration with structural break" *Oxford Bulletin of Economics and Statistics* **68**, 623-646.

Engle, R.F. and C.W.J. Granger (1987) "Co-integration and error correction: representation, estimation and testing" *Econometrica* **55**, 251-276.

Gregory, A. and B. Hansen (1996) "Residual-based tests for cointegration in models with regime shifts" *Journal of Econometrics* **70**, 99-126.

Juhl, T. and Z. Xiao (2009) "Tests for changing mean with monotonic power" *Journal of Econometrics* **148**, 14-24.

Kejriwal, M. (2008) "Cointegration with structural breaks: An application to the Feldstein-Horioka Puzzle" *Studies in Nonlinear Dynamics & Econometrics* **12**, 1-37.

Kejriwal, M. and P. Perron (2008) "The limit distribution of the estimates in cointegrated regression models with multiple structural changes" *Journal of Econometrics* **146**, 59-73.

Kejriwal, M. and P. Perron (2010) "Testing for multiple structural changes in cointegrated regression models" *Journal of Business and Economic Statistics* **28**, 503-522.

Kurozumi, E. (2002) "Testing for stationarity with a break" *Journal of Econometrics* **108**, 63-99.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Y. Shin (1992) "Testing the null of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of Econometrics* **54**, 159-178.

Perron, P. (2006) "Dealing with Structural Breaks" In: Patterson, K., Mills, T.C. (Eds.) Palgrave Handbook of Econometrics, 278-352.

Schwert, G. W. (1989) "Tests for unit roots: A Monte Carlo investigation" *Journal of Business and Economic Statistics* **7**, 147-159.