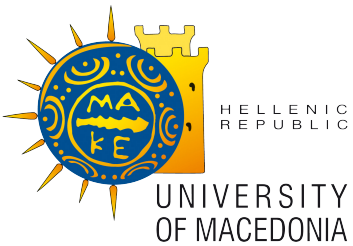
Bootstrap Cointegration Testing

Author: Stylianos Daskalinas
Supervisor: Theodore Panagiotidis

Masters Thesis

Interdepartmental Postgraduate Program in Economic Science
University of Macedonia

March, 2024



Abstract

This work assesses the power and size properties of the wild bootstrap implementation of Johansen's test for cointegration for a variety of realistic data generating processes using simulations on systems with random coefficients. A VAR(2) is used, with parameters generated randomly (but under some restrictions). We construct innovations processes with GARCH(1,1), EGARCH(1,1) and SARV(1,1) variances, as well as with outliers and regime shifts in the variance. We find that the wild bootstrap test significantly outperforms Johansen's test when innovations exhibit conditional heteroskedasticity or stochastic volatility but not when innovations have outliers or regime shifts. Also, we find a significant tendency for both tests to underestimate the cointegrating rank, which somewhat contradicts current literature (probably associated to our choice of intervals from which parameters are drawn).

Key words
Cointegration, Vector Autoregression, Johansen's test, Bootstrap, Resampling, Size, Power, Statistical Test Performance, Conditional Heteroskedasticity, Outliers, Stochastic Volatility, Structural Breaks

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Theodore Panagiotdis for providing insightful comments, useful guidance and advice, and multiple educational and professional opportunities, regarding both the present thesis and my academic course in general. His expertise in Econometrics greatly enriched this work. I would also like to thank the post-doc researcher, Georgios Papapanagiotou, for his exceptional assistance with R coding and simulations. Additionally, I am deeply thankful to the entire community of the Interdepartmental Postgraduate Program in Economics of the University of Macedonia. The involved professors and PhD students, and of course, my fellow postgraduate students, provided a friendly, competitive, and intellectually stimulating environment that motivated me even more to pursue an academic career. Finally, I owe a debt of gratitude to my parents, Yiannis and Eumorfia, and my brother Vassilis, for their unconditional support and encouragement throughout my postgraduate studies and the writing of this thesis.

Contents

List of Tables

List of Figures

Notation

Symbol	Description
p	Number of variables/dimension of a system
T	Number of realisations of a series/system
r	Cointegrating rank of a system
VECM(k)	Vector Error Correction model of lag length k
VAR(k)	Vector Autoregression/Autoregressive model of lag length k
$B(\nu)$	Standard Brownian motion (defined on $C[0,1]$)
Δ	First difference operator

Table 1: Notation

1 Introduction

1.1 Motivation & Contributions

In recent years, cointegration has become a most important tool for applied econometrics, especially in macroeconomic and financial applications. Despite the multitude of both single and multi-equation tests for cointegrating relations found throughout the literature, the Engle and Granger (1987) two step method and Johansen (1988, 1991) tests have dominated the applied landscape. An inherent advantage of Johansen's approach is the capability of identifying all possible cointegrating relations without speculating causal relations between variables. On the other hand, Johansen's approach produces two test statistics which converge to Brownian motion functionals, and therefore follow non-standard distributions. Thus proper critical values for Johansen's tests can be obtained through Monte Carlo methods. However, complicated Data Generating Processes[1] can introduce unknown parameters in Johansen's statistics, thus making them non-pivotal as they would depend on unknown quantities. A case where non-pivatility is likely to occur is when the innovations process driving a cointegrating system has special traits, such as conditional heteroskedasticity, which effectively introduces parameters and reshapes the specification of the model. Given that many real world time series do exhibit such traits, assessing their impact on inference procedures is important since their validity can make the difference between drawing insightful inferences and producing spurious results.

Bootstrapping the statistics is an effective method for bypassing non-pivotalities and certain cases of misspecification, as well as for handling small-sample cases, where the asymptotic results of Johansen's tests are challenged. The presence of dependent data following underlying processes slightly complicates producing bootstrapped versions of standard statistics, as is typically the case with dependent data. However, the literature has so far managed to pull through and provide bootstrapped versions of Johansen's test for cointegration. Its popularity with applied economists, as well as the fact that the testing procedure is straightforwardly based on the likelihood ratio format of statistical testing of nested models, are probably driving factors that motivated researchers to undertake bootstrapping this specific test(s) instead of other univariate/multivariate tests for cointegration.

The most well-known bootstrap implementations of Johansen's test(s) are the block bootstrap method (Giersbergen (1996)), the IID bootstrap method (Swensen (2006)) and the wild bootstrap method (Cavaliere et al. (2010a,b)). Each test has its merits and drawbacks, and each is fit for tackling different forms of issues that might arise in a cointegrated system. The block bootstrap mitigates autocorrelation and the presence of MA terms in a VECM, the IID bootstrap can provide robustness and bias reduction

[1] Henceforth abbreviated to DGPs.

against specific observations (much like the jackknife which it was inspired from, see Efron (1992)), and the wild bootstrap allows for overcoming heteroskedasticity, and time-dependent variance in general. Therefore, bootstrap methods can be used to resolve a wide variety of issues that could violate the assumptions of Johansen's test or compromise its effectiveness.

Of course, these methods have theoretical justifications and proofs verifying their limiting validity, however, putting these methods to the test is a necessary step before widely adopting them. As is usual with statistical hypothesis tests, we can evaluate the performance of bootstrap cointegration testing through simulations. There, we can run sort of "controlled experiments" where we can modify the DGPs accordingly to examine the robustness of these methods against the issues they are designed to resolve. Given that cointegration is typically applied in macroeconomic and financial time series, we consider features such systems' innovations processes are likely to exhibit. More specifically, we examine the power and size of bootstrap tests against:

- Conditional heteroskedasticity
- Stochastic volatility
- Outliers in variance
- Regime shifts in variance

The first two cases are a perfect match for the wild bootstrap test, while the last two are likely to pose a challenge, even though they concern anomalies in the variance of the model, while its deterministic specification remains intact. Each of these cases typically corresponds to real world situations and series such as:

- Risky assets prices
- Macroeconomic control variables (consumption, employment etc.) during periods of high uncertainty
- Price levels of countries that enter a monetary union
- Exchange rates after critical elections
- Agricultural commodity prices during natural disasters
- Companies' stock prices when launching a new product/service
- CDS spreads during financial crises

In the present work, we aspire to make three distinct contributions: First, we aim at verifying the findings of the existing literature regarding the performance (power and size) of cointegration testing and bootstrap cointegration testing methods, with a focus on the

wild bootstrap implementation of Johansen's test. Second, we attempt to extend the reach of simulations-based assessments of Johansen's and wild bootstrap tests by considering DGPs with random coefficients. This approach, despite being much more computationally demanding than the fixed coefficients approach, allows us to address not a specific model but a class of models, which expands the universality of any results regarding the performance of the statistical methods we consider. Third, we wish to examine whether the wild bootstrap, despite its structure and purpose, can actually alleviate any issues past time-dependent variance, since real world time series may exhibit more that one anomalies that must be accounted for.

1.2 Organisation

In this work our focus is inference on the number of long-run relations between a collection of variables (with a unit root), therefore, albeit important and useful, estimation of the models' coefficients, as well as inference on these coefficients, will not be discussed here. Also, we restrain our study to models with no deterministic components, for reasons that will be discussed in section 4.

Our study is organised as follows. First, we trek through the relevant literature and discuss landmark publications and important results, with a focus on simulation results. Then, we review Johansen's test and discuss where and how issues of non-pivotality may manifest. We also discuss the three major bootstrap implementations of Johansen's trace test, their procedures and properties. An extensive exposition of our methodology follows, including data generation procedures we test performance for, testing procedures algorithms, as well as computational details. Then we move to discuss our results. Ultimately, we summarize and consider directions for further studies.

2 Literature

Our discussion taps into the intersection of multiple parts of the cointegration analysis literature, namely, the bootstrap cointegration testing literature and the non-pivotal cointegration literature, as well as the simulations-based evaluation of statistical tests' performance.

2.1 Cointegration Testing

Besides the test introduced in Johansen (1988) which is probably the most frequently used for studies involving more than two series (where where it is possible for more than one cointegrating relations to exist), a great variety of methodologies have been constructed to assess the presence and number of long-run relations between random walks, as well as estimate the relations themselves. Pioneering contributions by Granger (1981) and Engle and Granger (1987) established a straightforward testing procedure based on retrieving the residuals of a regression among the collection of variables under investigation and testing for a unit root. Another single-equation test proposed by Shin (1994) uses an unobserved components setting to test whether the residual series has a random walk component, similar to the methodology implemented in Kwiatkowski et al. (1992). Tests based on residuals of linear models between unit root variables (both single and multi-equation) where proposed in Phillips and Ouliaris (1990) to trace the presence of cointegrating relations.

An early multivariate testing procedure proposed by Stock and Watson (1988) uses the common trends representation (roughly the "number of unit roots", which is equal to the dimension of the model minus the cointegrating rank) and principal components analysis. The systems testing procedure introduced in Johansen (1988) and elaborated in Johansen (1991) uses sequential likelihood ratio testing of nested reduced-rank-specified vector autoregressions in error correction form. Another multivariate test which determines the number of common trends is proposed by Nyblom and Harvey (2000), based on a variance ratio scheme. A non-parametric common trends test, based on variance ratios is derived in Breitung (2002). Cheng and Phillips (2009) propose estimating the number of cointegrating relations by comparing information criteria scores of different cointegrating rank specifications on a model similar to that of Johansen's. Later, they put their method to the test against systems with time-varying variance (Cheng and Phillips (2012)).

Hansen (1992) derived asymptotics on a single-equation cointegration model with the errors' standard deviation modelled as a random walk. He demonstrated that a large part of standard single-equation asymptotics for cointegration remain unchanged under time-varying (and possibly unbounded) variance. Using simulations of a bivariate cointegrated fixed coefficients system with GARCH(1,1) and EGARCH(1,1)-varying errors and

a variety of cointegration tests, Lee and Tse (1996) concluded that the size and power of Johansen's tests is not significantly affected by the presence of heteroskedasticity in the error terms, however the size increases as the GARCH coefficients' sum reaches 1, especially in small samples. Höglund and Östermark (2003) extended the discussion by adding structural breaks in the simulations' DGPs and examining the size and power of various cointegration tests. Using a bivariate DGP with one lag and fixed coefficients, they concluded that single-equation tests perform better and face lower size distortions. Additional simulations on a three-variable system of fixed coefficients by Silvapulle and Podivinsky (2000) showed that ARCH and GARCH-varying innovations are likely to challenge the robustness of Johansen's tests as the variance processes approximate a unit root shape. Also, they demonstrated the sensitivity of Johansen's test when innovations come from different distributions (other than the normal). The combination of non-normal distributions and ARCH/GARCH innovations led to significant size distortions. Maki (2013) runs simulations on a bivariate system of independent series (true rank is always 0) with MGARCH, FIGARCH and BEKK innovations, as well as variance breaks, for a variety of both single and multi-equation tests. Conditional heteroskedasticity is found to affect the power of Johansen's maximum eigenvalue test, and the author proposes using the Breitung (2002) test instead when data are likely to exhibit conditional heteroskedasticity.

Paper	p	Model	Innovations	T	Reps
Lee and Tse (1996)	2	VAR(1)	GARCH(1,1), EGARCH(1,1)	100, 1,000	10,000
Haug (1996)	2	VAR(2)	ARCH(1)	40, 100	5,000
Silvapulle and Podivinsky (2000)	3	VAR(1) with intercept	ARCH(2), GARCH(1,1)	50, 100, 500	20,000
Maki (2013)	2	VAR(1)	MGARCH, FIGARCH, BEKK, break	200, 400	10,000

Table 2: Current literature cointegration testing simulations with p variables, T realisations

2.2 Bootstrap Cointegration Testing

An early bootstrap implementation of Johansen's trace test was developed by Giersbergen (1996), where, for bootstrap sample generation, a random-length block sampling scheme (over estimated residuals) was proposed, aiming to account for power loss due to correlation between adjacent innovations. The test statistic's distribution was approximated through the empirical CDF of the bootstrap sample. Using simulations with two bivariate DGPs (with

and without MA terms), he demonstrated a remarkable success rate in correctly accepting and rejecting the null of no cointegration when pseudo-samples are generated using small blocks (close to an IID sampling case).

The performance of an IID sampling bootstrap version of both Johansen's trace and maximum eigenvalue tests in small samples was discussed through simulations in Harris and Judge (1998). Using a three-variable constant coefficient DGP, the authors demonstrated the capacity of bootstrap-based testing in correctly rejecting the null, however they noticed a power loss with cases of large cointegrating rank. Swensen (2006) provided proof for the consistency of the IID bootstrap in approximating the actual distribution of Johansen's trace statistic when the roots of the characteristic polynomial are either equal to one or lie outside the unit circle. This condition effectively means that the systems must exhibit neither explosiveness nor seasonality. Leveraging this condition (which allowed invoking other results such as the Granger representation theorem) along with some additional conditions (innovations with zero mean and constant variance matrix), he has managed to provide asymptotics for the convergence of the IID bootstrap empirical CDF. Swensen reinforced the asymptotic results with simulations on a four-variable fixed coefficients DGP.

Cavaliere and Taylor (2006) studied an unobserved components model with the null of stationary errors, based on the test by Shin (1994), with a time-varying "system"-variance matrix exhibiting structural breaks. They provided proof of the non-pivotality of Shin's statistic under this variance scheme and proposed a bootstrap version of the statistic to resolve that. Cavaliere et al. (2010a) considered a Johansen-based cointegration model with non-stationary unconditional volatility and proved that Johansen's tests are distributed as volatility matrix-dependent martingale functionals, and therefore are non-pivotal. Authors proposed a wild bootstrap scheme for estimating the trace test's distribution and obtaining p-values, using similar conditions regarding the roots of the characteristic equation of the VECM as those in Swensen (2006), in order to take advantage of their convenient asymptotic consequences. They proved the validity of the wild bootstrap and its robustness against a generic volatility process. The authors put their method to the test with a 5-variable fixed coefficients model without lags and at most one cointegrating relation ++++

A similar model with stationary conditionaly heterskedastic innovations is discussed in Cavaliere et al. (2010b), where it is proven that Johansen's statistic is asymptotically pivotal under such innovations processes. After properly formatting assumptions to allow for the wild bootstrap to be implemented and for asymptotics similar to Swensen (2006) to hold, the authors put their model to the test with a 5-variable fixed coefficients VAR(1) specification and a variety of innovations processes (GARCH, EGARCH etc.). The wild bootstrap test is found to significantly alleviate size distortions. The wild bootstrap

cointegration test is also discussed in Cavaliere et al. (2012), where additional asymptotics on the validity of the method, as well as its asymptotic correctness are provided. The authors reinforce their results with simulations on a 4-variable fixed coefficients VAR(2). Cavaliere et al. (2015) compare the statistical properties and performance of sequential inference and information criteria inference for cointegrated VECMs with conditionally heteroskedastic innovations. They run simulations with a VAR(2) model with rank no more than 2 and innovations with GARCH, SARV and regime shift variance, as well as IID innovations. They find that the Bayesian information criterion and wild bootstrap test have outperform the rest.

Paper	p	Model	Innovations	T	B	Reps
Giersbergen (1996)	2	VAR(1), VAR(2), VARMA(2,1)	IID	100	99	1,000
Harris and Judge (1998)	3	VAR(1)	IID	50	5,000	5,000
Swensen (2006)	5	VAR(1)	IID	100	5,000	10,000
Cavaliere et al. (2010a)	5	VAR(1)	random break	100, 200, 400	399	10,000
Cavaliere et al. (2010b)	5	VAR(1)	GARCH, EGARCH,GJR-GARCH	100, 200, 400	399	10,000
Cavaliere et al. (2012)	4	VAR(2)	IID	50, 100, 200	399	10,000
Cavaliere et al. (2015)	4	VAR(2)	GARCH, break, EGARCH, SARV, IID	50, 100, 200, 400	399	10,000

Table 3: Current literature bootstrap cointegration testing simulations with p variables, T realisations

3 Theoretical Framework

3.1 Review of Cointegration

Consider a collection of p time series $\{Y_t = (Y_{1t}, \ldots, Y_{pt})'\}_{t=1\ldots T}$, with $Y_t \sim I(1)$, i.e. each series has a unit root. A cointegrating relation between $Y_{1t}, \ldots, Y_{pt}$ is a $p \times 1$ vector β such that $\beta' Y_t \sim I(0)$. Intuitively, the presence of a cointegrating relation among a collection of $I(1)$ variables is a linear combination that keeps non-stationary series from drifting too far apart over time.

Such relationships are often found in macroeconomic and financial series, where various factors might influence different series but they are still linked in the long term by some underlying equilibrium which corresponds to some sort of economic, financial or structural relationship between them. Such relations can be used to make informed policy or investment decisions. Examples include evaluating the possibility of a successful monetary union (Haug et al. (2000)), assessing the optimal taxation strategy for energy-related goods (Vücel and Guo (1994)), and finding suitable stocks to apply pairs trading strategies (Caldeira and Moura (2013)).

It is possible for up to $p-1$ such relations to exist (p relations would imply Y_t is the sum of stationary series, hence also stationary). Therefore, in general, a full-rank $p \times r$ matrix β holds r cointegrating relations if $\beta' Y_t = (U_{1t}, \ldots, U_{rt}) \sim I(0)$.

3.2 Johansen's test

Johansen (1988) introduced a procedure for inferring r through reduced rank ML estimation of a VECM model. Assume a VECM(k) of the form

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{i=1}^{k} \Gamma_i \Delta Y_{t-i} + \Phi D_t + \epsilon_t \tag{1}$$

Here, the innovations terms are assumed to be Gaussian $\epsilon_t \sim NIID(0, \Omega)$ but the model may be generalised to a variety of zero-mean stationary processes (see Boswijk et al. (2015)). Also, any deterministic terms are suppressed (intercept, trend, dummies etc.) in D_t. Johansen's method depends on assuming a reduced rank specification, i.e. $H_r : \Pi = \alpha_r \beta_r'$, where the full-rank $p \times r$ matrices β_r and α_r hold cointegrating vectors and long-run adjustment parameters respectively. Two tests proposed by Johansen have the alternatives $H_{r+1} : \Pi = \alpha_{r+1} \beta_{r+1}'$ and $H_p : \Pi = \alpha_p \beta_p'$, the so called maximum eigenvalue and trace tests respectively. We shall focus on the trace test, aligning with most of the bootstrap cointegration testing literature.

Inference on the rank of α_r and β_r is done through sequential likelihood ratio testing, starting with $r = 0$, where Π is a $p \times p$ matrix of zeros. Under the null, the model is reduced to its concentrated form $R_{0t} = \alpha\beta' R_{1t} + \epsilon_t$, where R_{0t} and R_{1t} are the residuals of regressing ΔY_t and Y_{t-1} to $\Delta Y_{t-1}, \ldots, \Delta Y_{t-k}, D_t$. Maximizing the likelihood function under the null, first with respect to α and Ω and then with respect to β yields:

$$(L_r^*)^{-T/2} = |S_{00}| \prod_{i=1}^{r} (1 - \lambda_i) \tag{2}$$

Here $\lambda_1, \ldots, \lambda_r$ are the r largest eigenvalues of the problem $|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0$ with $S_{ij} = \frac{1}{T} \sum_{t=k+1}^{T} R_{it} R'_{jt},\ i, j = 0, 1$. Thus, we can easily construct the likelihood ratio of every null. The trace test is formulated as the difference between the maximized log-likelihoods of the two hypotheses, which with the eigenvalues ordered as $\lambda_1 \geq \cdots \geq \lambda_p$ becomes:

$$-2 \log(LR(H(r)|H(p))) = -T \sum_{j=r+1}^{p} \log(1 - \lambda_j) \tag{3}$$

Therefore, the test statistic will involve all of the eigenvalues if $H_0 : r = 0$, all but the first if $H_0 : r = 1$, and so on, until $H_0 : r = p - 1$, where only the last eigenvalue will be involved. This is directly related to the nested nature of cointegrating relations (the presence of $r+1$ cointegrating relations implies the presence of r cointegrating relations).

As is evident, we need the eigenvalues of the model to lie within the $[0, 1)$ interval in order for the statistic to not diverge to infinity or take negative values. As usual with one-sided hypothesis testing, we reject the null if the value of the test statistic exceeds the corresponding critical value. Unfortunately, the trace test statistic's asymptotic distribution converges to the rather complicated Brownian motion functional:

$$-2 \log(LR(H(r)|H(p))) \xrightarrow{\text{w}} tr\left\{ \int_0^1 dB F' \left(\int_0^1 F F' \right)^{-1} \int_0^1 F dB' \right\} \tag{4}$$

Here, $B(u)$ is a $p - r$ standard Brownian motion defined on $C[0, 1]$ and $F(u)$ is a functional that depends on B and the deterministic specification (constant, trend etc.) of the model. In our case, where deterministic terms are neglected to allow for the generation of suitable systems in a computationally reasonable timespan, F is simply defined as $F_i(u) = B_i(u)$, with $i = 1, \ldots, p - r$. It is worth noting that two different types of misspecification will lead to issues in two different parts of the inference procedure. Estimation with the wrong number of lags will result in incorrect eigenvalues, while incorrect specification of the deterministic term will lead to the use of improper critical values.

The limiting distribution of the trace test is a non-standard one, therefore critical

values can only be obtained through Monte Carlo simulations. A table of critical values is given in Johansen (1995), while improved approximations are available in MacKinnon et al. (1999). Nevertheless, the lack of definitive critical values, along with issues of misspecification and non-pivotality do invite the use of alternative methods such as the bootstrap to assess the cointegrating rank.

3.3 Bootstrap Implementations of Johansen's Test

As is usual in bootstrap-based inference, the aim of all bootstrap implementations of Johansen's trace test is to attempt to construct the sampling distribution of the test statistic out of the given sample, by producing a bootstrap sample of the statistic. This method typically depends on calculating the value of the statistic for a number of pseudo-samples, produced with replacement sampling or some other method, out of the original data.

Besides IID sampling, there are a few bootstrap methodologies suitable for dependent data, including the wild bootstrap (Cavaliere et al. (2020) provide an extensive discussion for single series cases) and the block bootstrap. The core idea of bootstrap implementations of parametric statistical tests in time series models can be summed up in the following algorithm:

Steps
Estimate the parameters and residuals of the relevant model.
Repeat B times:
a. Use a suitable method to generate a pseudo-sample of residuals series.
b. Create a series pseudo-sample by recursively constructing the series using estimates and the residuals pseudo-sample.
c. Calculate and save the desired statistic of the series pseudo-sample.
Construct the empirical CDF of the statistic using the collected pseudo-sample statistics.
Calculate the p-value at the original sample's statistic using the empirical CDF.

Table 4: Parametric bootstrap inference process

As B approaches infinity, the empirical CDF should approximate the CDF of the statistic. This procedure does not depend on critical values, therefore inference with time series (where statistics typically converge to Brownian motion functionals) can simply be done by comparing the calculated p-values with the desired significance level. However, the problem transfers from critical values to proper sampling implementations and computational

feasibility. One should keep in mind that any results regarding the sampling distribution of the statistic of interest are relevant to the sample only, and discussing the distribution of the statistic for the whole population requires that we actually have a representative sample.

Regarding Johansen's test, the sequential testing procedure remains exactly the same, but with p-values and significance levels instead of statistics and critical values. With the bootstrap approach, the presence of deterministic terms does not affect the procedure, unlike the actual critical values of the test, where the functional $F(\nu)$ takes different forms depending on the structure of the deterministic terms. The same goes for many cases of non-pivotality, since (given proper estimation) the asymptotic distribution of the statistic will not affect inference.

However, one should proceed with caution as incorrect specification does translate to incorrect trace statistic evaluation which is possible to compromise bootstrap-based inference. Our discussion focuses on the three best known bootstrap implementations of Johansen's test, the block bootstrap of Giersbergen (1996), the IID bootstrap of Swensen (2006) and the wild bootstrap of Cavaliere et al. (2010a).

3.3.1 Block Bootstrap

The bootstrap implementation introduced in Giersbergen (1996) depends on constructing pseudo-sample innovations through binding blocks of adjacent estimated innovations. This ensures that despite building the innovations out of random sampling, the structure of adjacent innovations will be maintained, and therefore any effects involved will also manifest in the bootstrap pseudo-samples. Due to this advantage, block-based sampling is quite frequent in statistics with dependent data.

Giersbergen suggests the use of a moving, random-length block sampling. Each block's starting point is randomly drawn from the discrete uniform distribution, while the block's length is randomly chosen from the geometric distribution, with the distribution's parameter value (probability of "success") decided by the researcher (it is found that smaller blocks function better in correctly rejecting the null). Once enough blocks are sampled to build T innovations, the innovation pseudo-sample can be used to recursively construct the pseudo-sample series and estimate the trace statistic for the replication.

Once a sufficient number of replications is assembled, the null should be rejected if the trace statistic for the actual sample exceeds the $1-a$ quantile[2] of the empirical cumulative distribution of the pseudo-sample statistics.

[2] a being the chosen statistical significance level.

The algorithm for Giersbergen's block bootstrap implementation is:

Steps
Calculate estimates $\alpha, \beta, \Phi, \Gamma_1, \ldots, \Gamma_k, \{\hat{\epsilon}_t\}_{t=1}^T$ at null r.
For $b = 1, \ldots, B$
a. Sample blocks of random length (geometric) and starting point (uniform) from $\{\hat{\epsilon}_t\}_{t=1}^T$ until a pseudo sample $(\{\hat{\epsilon}_t^*\}_{t=1}^T)_b$ of T realisations is amassed.
b. Generate series pseudo-sample $(\{Y_t^*\}_{t=1}^T)_b$ recursively from $\Delta Y_t^* = \hat{\alpha}\hat{\beta}' Y_{t-1}^* + \sum_{i=1}^k \hat{\Gamma}_i \Delta Y_{t-i}^* + \hat{\Phi} D_t + \hat{\epsilon}_{b,t}^*$.
c. Calculate and save the bootstrap realisation trace statistic LR_b^* with $(\{Y_t^*\}_{t=1}^T)_b$.
Construct the empirical CDF of the trace statistic using pseudo-samples values.
Calculate the p value for the null of cointegrating rank r at the original sample statistic using the empirical CDF as $p = 1 - \hat{G}_B(LR)$.

Table 5: Giersbergen (1996) block bootstrap test algorithm

Despite yielding promising results, this procedure has not yet been tested against more demanding DGPs than the constant variance IID innovations cases and, as far as the author has searched, there is no decisive mathematical proof of consistency in large samples (although for unit-length blocks, Swensen's proof may also cover this sampling scheme).

3.3.2 IID Bootstrap

The main point of the IID bootstrap is that pseudo-samples of innovations are constructed by picking out innovation estimates using independent draws from the uniform distribution over the number of realisations. As was discussed in the literature section, using random sampling for recreating a sample was initially examined by Harris and Judge (1998), with no definitive proof of the consistency of the statistic.

A more rigorous treatment of the IID bootstrap, as well as the proof of its validity in approximating the cumulative distribution of the trace statistic was provided in Swensen (2006). Swensen focused on systems where the roots of the characteristic equation of the VECM, $det(A(z)) = 0$, with $A(z) = (1-z)I_p - \alpha\beta' z - \sum_{i=1}^k \Gamma_i (1-z) z^i$ are either equal to 1 or outside the unit circle. This excludes explosiveness (roots in the unit circle) and seasonality (roots equal to -1). Also, together with the condition that $det(\alpha_\perp'(I_p - \sum_{i=1}^k \Gamma_i)\beta_\perp) \neq 0$, these two ensure that the Granger representation theorem (as given in Johansen (1995)) holds, which allows expressing the cointegrated system as a collection of sums of innovation terms. This opens the model and its statistics up to many of the asymptotic results around partial sums.

Swensen's approach does not significantly depart from the initial intention of the bootstrap method, which is to reduce bias introduced by specific observations (Efron (1992)). Despite not being designed with dependent data in mind, like the wild or block bootstrap methods, the IID bootstrap is most likely effective in addressing a variety of specification and small sample issues, and should provide some level of robustness against tail events such as outliers.

The algorithm proposed by Swensen depends on generating bootstrap samples of innovations through random sampling with replacement of the centered estimated residuals series.

Steps
Calculate estimates $\hat{\Phi}, \hat{\Gamma_1}, \ldots, \hat{\Gamma_k}$ at null p and estimates $\hat{\alpha}, \hat{\beta}$ and $\{\hat{\epsilon}_t\}_{t=1}^T$ at null r.
Ensure $det(\hat{A}_{r^*}(z)) = 0 \Rightarrow z \in \mathbb{R} \setminus [-1, 1)$ and $det(\alpha'_\perp(I_p - \sum_{i=1}^k \Gamma_i)\beta_\perp) \neq 0$
For $b = 1, \ldots B$
a. Sample randomly from centered innovations $\{\hat{\epsilon}_t - \frac{1}{T}\sum_{j=1}^T \hat{\epsilon}_j\}_{t=1}^T$ to build a bootstrap pseudo-sample of innovations, $(\{\hat{\epsilon}_t^*\}_{t=1}^T)_b$.
b. Generate series pseudo-sample $\{Y_t^*\}_{t=1}^T$ recursively from $\Delta Y_t^* = \hat{\alpha}\hat{\beta}' Y_{t-1}^* + \sum_{i=1}^k \hat{\Gamma}_i \Delta Y_{t-i}^* + \hat{\Phi} D_t + \hat{\epsilon}_{b,t}^*$, with actual series values as initial values.
c. Calculate and save the bootstrap realisation trace statistic LR_b^* with $(\{Y_t^*\}_{t=1}^T)_b$.
Construct the empirical CDF of the trace statistic, $\hat{G}_B(\cdot)$, using pseudo-samples values $\{LR_b^*\}_{b=1}^B$.
Calculate the p value for the null of cointegrating rank r at the original sample statistic using the empirical CDF as $p = 1 - \hat{G}_B(LR)$.

Table 6: Swensen (2006) IID bootstrap test algorithm

This procedure, despite being quite successful in estimating the cointegrating rank, is somewhat restricting, since it is inapplicable to systems the (estimated) characteristic equations of which yield roots that either lie inside the unit circle (explosiveness) or are equal to -1 (seasonality).

3.3.3 Wild Bootstrap

The wild bootstrap implementation is somewhat different than the IID case, since generating pseudo-samples does not depend on resampling but on multiplication with another series, typically scalar, IID with mean 0 and variance 1. Unlike the IID bootstrap, WB only requires that the innovations process is a martingale difference sequence with respect to past data, with finite 4^{th} order moments and the variance process converging to a constant

matrix, establishing a sort of global homoskedasticity. The wild bootstrap also requires that the roots of the characteristic equation are not inside of the unit circle or equal to -1, which, along with a few more conditions, suffices to provide room for partial sums asymptotics, much like Swensen's test. The main advantage of the wild bootstrap method for pseudo-sample generation is that the ordering of the innovations, and consequently the underlying structure of the (time-dependent) variance process, are directly relayed into the bootstrap innovations pseudo-sample, without any violations of the assumptions constituting the innovations process of the model. Also, it is worth noting that the actual variance process of the initial model is also relayed into the pseudo-samples due to the specification of the pseudo-innovations.

The following algorithm, introduced in Cavaliere et al. (2010a,b) encapsulates the procedure of the wild bootstrap implementation of Johansen's test.

Steps
Calculate estimates $\hat{\Phi}, \hat{\Gamma_1}, \ldots, \hat{\Gamma_k}$ at null p and estimates $\hat{\alpha}, \hat{\beta}$ and $\{\hat{\epsilon}_t\}_{t=1}^T$ at null r.
Ensure $det(\hat{A}_{r^*}(z)) = 0 \Rightarrow z \in \mathbb{R} \setminus [-1, 1)$ and $det(\alpha'_\perp(I_p - \sum_{i=1}^k \Gamma_i)\beta_\perp) \neq 0$
For $b = 1, \ldots B$
a. Build a bootstrap pseudo-sample of innovations, $(\{\hat{\epsilon}_t^*\}_{t=1}^T)_b$ as $\hat{\epsilon}_{b,t}^* = \hat{\epsilon}_t w_t$, where $w_t \sim N(0,1)$
b. Generate series pseudo-sample $\{Y_t^*\}_{t=1}^T$ recursively from $\Delta Y_t^* = \hat{\alpha}\hat{\beta}' Y_{t-1}^* + \sum_{i=1}^k \hat{\Gamma}_i \Delta Y_{t-i}^* + \hat{\Phi} D_t + \hat{\epsilon}_{b,t}^*$, with actual series values as initial values.
c. Calculate and save the bootstrap realisation trace statistic LR_b^* with $(\{Y_t^*\}_{t=1}^T)_b$.
Construct the empirical CDF of the trace statistic, $\hat{G}_B(\cdot)$, using pseudo-samples values $\{LR_b^*\}_{b=1}^B$.
Calculate the p value for the null of cointegrating rank r at the original sample statistic using the empirical CDF as $p = 1 - \hat{G}_B(LR)$.

Table 7: Cavaliere et al. (2010a,b) wild bootstrap test algorithm

The wild bootstrap is the only method capable of accounting for time-dependent variance in the VECM's innovations, which is a trait met most frequently (if not almost certainly) in financial and macroeconomic time series. The drawback of the method is that it does not account for observation-specific bias and therefore has robustness limitations against tail events.

4 Methodology

4.1 Data Generating Process

For our experiments, we follow Cavaliere et al. (2012), Cavaliere et al. (2014) and Cavaliere et al. (2015), and implement the following DGP:

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \Gamma \Delta Y_{t-1} + \epsilon_t \tag{5}$$

This DGP can address a wide variety of economic systems as it contains a relatively versatile structure with respect to cointegrating rank and also contains a lagged term, thus introducing memory in the process. The innovation term ϵ_t is key for inducing "complications" which match real world time series, such as heteroskedasticity or stochastic volatility.

4.1.1 Coefficients

Unfortunately, implementing this DGP directly (i.e. with randomly drawn matrices) can be quite difficult, since large or inappropriate values of the α, β and Γ matrices are likely to produce explosive or stationary systems, therefore most studies fix either some or all the values of α, β and Γ. In this study we make an attempt towards additional robustness through covering a large number of cases by ensuring that all of the coefficients, while still subjected to some restrictions, remain randomly generated, therefore each replication corresponds to a different system where only properties concerning the cointegrating rank and innovations remain the same. Unfortunately, even with the application of restrictions, the possibility of producing explosive or stationary series still remains, therefore we look for properties that will minimize the number of discarded systems.

After running various experiments with combinations of constraints on the matrices we have arrived at some relatively functional DGP constraints:

1. α and β composed of the first r rows of two different orthonormal matrices.
2. α, β, composed of the first r rows of the same orthonormal matrix (and thus equal).
3. α, β, composed of the first r rows of the same orthonormal matrix but with perturbations applied to β (by adding uniform draws from small intervals) so that the two matrices are not equal.
4. α composed of uniform random draws from a small interval (say the $[-0.3, 0.3]$ interval) and β composed of the first r rows of an orthonormal matrix.

These specifications minimize the number of discarded systems we need to go through to reach the desired number of replications for our simulations to bear any meaning. The

second specification has the best results in terms of time and computational feasibility, however it only yields symmetric VECMs, therefore its utility in addressing a wide variety of systems is limited. Hence, we choose the third specification which, despite not having the generality of the first one, still manages to produce asymmetric systems for our $10,000$ replications. It is worth noting that the presence of orthonormal matrices has significant effect in generating appropriate systems.

The following matrices addressing the $p = 3,\ r = 2$ case are indicative of the possible generality our suggested DGP can reach

$$\alpha = \begin{bmatrix} 0.7733767 & -0.6147086 \\ 0.5470272 & 0.7706482 \\ -0.3203899 & -0.1680319 \end{bmatrix} \qquad \beta = \begin{bmatrix} 1.0812370 & -1.1726272 \\ 1.0012929 & 0.4369205 \\ -0.1059243 & -0.5604168 \end{bmatrix}$$

Evidently, this method can reach any suitable combinations of values, however departing the $[-2, 2]$ interval is unlikely, at least without jeopardizing the success rate of generated systems. On the other hand, when using totally random coefficients (i.e. simply drawn from a uniform distribution), even after applying techniques to normalise the magnitude of the product's eigenvalues, generated systems only fall marginally within our requirements ($I(1)$ with specific cointegrating rank), and reaching $10,000$ valid systems can take even ten times longer. It is likely that the diagonal elements of the reverted product of (the first r rows of) an orthonormal matrix with itself are responsible for this improvement.

In the general reduced-rank VECM, the presence of lagged terms coefficients should only minimally affect cointegration and indeed, the values of the Γ parameter do not seem to disrupt the generation of systems with the desired cointegrating rank. However, they do seem to complicate the generation of strictly $I(1)$ series as large values tend to result in series with either a second unit root or in explosiveness (at least when verified with the augmented Dickey-Fuller test), therefore useful systems are much harder to come by and computation becomes infeasible. Ultimately, experimentation with multiple values has shown that explosiveness and additional unit roots are much more likely to avoid if the eigenvalues of Γ are of magnitude smaller than the unit (and preferably not close to it). Thus, the eigenvalues matrix is generated by drawing values from the uniform distribution over small intervals. The magnitude of the eigenvectors matrix seems to have less of an effect, therefore drawing from larger intervals is also feasible. The Γ matrix is simply constructed by composing eigenvalues and eigenvectors matrices.

4.2 Deterministic Terms

Deterministic terms such as an intercept or a trend polynomial are not unlikely to be met in a VECM and accounting for their presence is important. However, including deterministic components of non-negligible magnitude in the model has proven to produce manageable systems at a very slow rate. Therefore, we align with the majority of the cointegration testing simulations literature and do not include them in our experiments.

4.3 Innovations Processes

Now, let our concentration shifts towards the innovations process, which, except for the zero-mean and stationarity assumptions, can be modified to incorporate a variety of phenomena without compromising computational feasibility. First, we assume that the innovations series ϵ_t has a variance process Σ_t, and the specification

$$\epsilon_t = \sigma_t z_t \tag{6}$$

where σ_t are $p \times p$ matrices of non-negative elements such that $\sigma_t \sigma_t' = \Sigma_t$, and $z_t \sim N(0, I_p)$. Therefore the innovations process is linearly distributed, with zero mean and the desired variance. Also, the resulting innovation series ϵ_t is normally distributed, therefore no complications arise from using the Gaussian likelihood function. From here on, there is significant flexibility in defining σ_t so that our systems exhibit interesting traits. The trivial constant variance case would only require σ_t to be a matrix of positive constants such that the product $\sigma_t \sigma_t'$ is positive-definite.

4.3.1 Conditional Heteroskedasticity

The presence of conditional heteroskedasticity in the series' innovations effectively implies that not every innovation comes from the exact same distribution, nor are the values of current innovations independent of past innovations, therefore the IID assumption of Johansen's model is violated. In this case, σ_t would be a matrix comprised of (the square roots of) the variance and covariance series of each variable. Autoregressive conditional heteroskedasticity is a common trait of financial and macroeconomic time series and likely to manifest in high frequencies. Since the paper by Engle (1982), a large variety of specifications for conditional heteroskedasticity have emerged (some with the possibility for covariance between the series). In this work, we focus on the specifications introduced in Bollerslev (1986) and Nelson (1991). Thus we set σ_t to be a diagonal matrix with entries $\sqrt{h_{i,t}}$, where $h_{i,t}$ are defined as

$$h_{i,t} = \omega + \sum_{j=1}^{q_1} a_j \epsilon_{i,t-j}^2 + \sum_{j=1}^{q_2} b_j h_{i,t-j} \tag{7}$$

and

$$h_{i,t} = \exp\Big(\omega + \sum_{j=1}^{q_1}[a_j\theta_{i,t-j} + g_j(|\theta_{i,t-j}| - E|\theta_{i,t-j}|)] + \sum_{j=1}^{q_2} b_j \log(h_{i,t-j})\Big) \tag{8}$$

Typically, $\theta_{it} = \frac{\epsilon_{i,t}}{\sqrt{h_{i,t}}}$, and for Gaussian $\epsilon_{i,t}$, the term $E|\theta_{i,t}|$ converges to $\sqrt{\frac{\pi}{2}}$. The GARCH specification is a standard measure of conditional heteroskedasticity, while the EGARCH specification can be used to capture asymmetric effects of the innovations process. It is worth noting that when the sum of the coefficients of the GARCH model approaches the unit, the long-run variance of ϵ_t becomes near-singular, which is the case we focus on. The EGARCH model maintains stability for any real parameters. In order to maintain the generality promised in the coefficients sub-section, we draw parameters from the uniform distribution, a_i and b_i from a positive interval (for GARCH we fix intervals so that their sum is close to one) and g_i from a negative interval (so that negative excess innovations increase variance, much like what frequently holds for financial time series). We choose single lags specifications for both models of heteroskedasticity ($q_1 = q_2 = 1$) since, as pointed out in Malmsten and Teräsvirta (2010), they are the specifications most frequently met in the literature and because addition of further lag terms will, at least for the GARCH model, complicate the proper setting of parameters.

4.3.2 Stochastic Volatility

In order to implement stochastic volatility, we incorporate a variant of the SARV(1,1) model, similar to those of Asai et al. (2006) and Cavaliere et al. (2014). This model uses an exponentiated autoregressive structure to capture the time-varying nature of volatility in economic and financial time series data. Unlike the deterministic process observed in conditional heteroskedasticity, stochastic volatility reflects the inherent uncertainty and unpredictability of market dynamics.

The "standard deviation" process (σ_t) is represented as a diagonal matrix with entries given by the exponential of half of the volatility level ($h_{i,t}/2$), where $h_{i,t}$ follows the SARV(1,1) structure:

$$h_{i,t} = ah_{i,t-1} + bv_{i,t} \tag{9}$$

Here, a and b are parameters governing the persistence and impact of past and present volatility shocks, while $v_{i,t} \sim N(0,1)$ represents the innovation process driving volatility dynamics. Unlike traditional GARCH specifications, stochastic volatility models allow for a more flexible representation of volatility dynamics, capturing the complicated interactions between past and current shocks without imposing strong structural assumptions.

4.3.3 Outliers

Tail events, such as outliers, can significantly impact the dynamics of economic and financial time series data, potentially leading to distorted estimations and misleading conclusions in cointegration analysis (Nielsen (2004)). These anomalies can manifest in both the deterministic and stochastic components of the data generating process (DGP), with high-magnitude fluctuations occurring sporadically due to unexpected events or extreme market conditions. While outliers in the deterministic part of the model can be represented as dummies with unknown dates of high magnitude, we focus on inducing special features through the innovations process to ensure computational feasibility.

In our simulation framework, we introduce outliers into the variance-covariance matrix (Σ_t) of the innovations process. We adopt a specification where $\Sigma_t = CG_t$, with C representing a positive definite matrix of nonnegative entries and $G_t = I_p + \sum_{i=1}^{k} G_i 1_{\{t=t_i\}}$ being a matrix that augments the constant variance with high-magnitude perturbations at specific points in time. These perturbations are represented by positive definite matrices G_i, $i = 1, \ldots, k$, of nonnegative entries, applied at outlier occurrence dates $t_1, t_2, \ldots, t_k$. The choice of high-magnitude perturbations challenges the performance of the test procedures by introducing significant deviations from the standard variance structure.

Outliers are likely to occur in real-world macroeconomic and financial data, corresponding to rare events or sudden shocks that disrupt the underlying dynamics of the system. By incorporating outliers into our simulation model, we aim to capture these irregularities and assess their impact on the performance of the test pocedures we consider. To maintain realism, we consider a case of three outliers scattered randomly across the system's realizations, each acting as approximately $10\times$ magnifications of the variances and covariances of the variables.

4.3.4 Variance with Structural Breaks

Structural breaks, often associated with significant events or changes in economic conditions, can have a lasting impact on financial and macroeconomic time series data. These breaks signify a shift in the underlying dynamics of the system, potentially leading to changes in long-term relationships and volatility patterns. For our simulations, we consider the scenario where structural changes occur in the variance of the model, akin to the approach discussed in Cavaliere et al. (2006).

To incorporate structural breaks into the variance process of the innovations, we model the covariance matrix Σ_t as a sum of step functions representing regime shifts. Formally, we

express Σ_t as:

$$\Sigma_t = \Omega_0 + \sum_{i=1}^{k} 1_{\{t \in T_i\}} \Big(\Omega_i - \Omega_0 \Big) \tag{10}$$

Here, T_i, $i = 1, \ldots, k$ denotes the periods during which the regime shifts occur, with Ω_i representing the variance levels associated with each regime. To ensure the positive definiteness of Σ_t, we ensure that the periods T_i do not intersect.

In our simulations, we focus on the case of a single structural break, randomly occurring past the first 30 and before the last 30 realizations. We restrict the period the break can occur in in order to ensure two distict "eras" within each sample and avoid cases where the break will act as an outlier affecting a few realisations sequentially. The break is specified as an increase in variance by k standard deviation units, with k randomly drawn from the interval $[1.5, 2.5]$. This setup mimics the potential impact of a significant event, such as a stock market collapse or a natural disaster, leading to a permanent negative effect on the system.

4.4 Simulations Methodology & Structure

Overall, our DGP for cointegrating rank r can be expressed as

$$\begin{cases} \Delta Y_t = A_r(A_r + R_r)' Y_{t-1} + D\Lambda D^{-1} \Delta Y_{t-1} + \epsilon_t \\ \epsilon_t = \sigma_t z_t \end{cases} \tag{11}$$

where A_r is a $p \times r$ matrix composed of the first r rows of an orthonormal matrix, R_r is a $p \times r$ matrix with random draws from the $[-0.7, 0.7]$ interval. D is a non-singular $p \times p$ matrix and Λ is a $p \times p$ diagonal matrix, filled with elements drawn uniformly from the $[-1, 1]$ and $[-0.5, 0.5]$ intervals accordingly (simulations also produce satisfactory results in reasonable time with R_r, D and Λ constructed from other subsets of the $[-1, 1]$ interval). Special features such as conditonal heteroskedasticity are introduced in the model from the structure of σ_t. For the cases of no cointegration ($r = 0$), we simply use the model $\Delta Y_t = D\Lambda D^{-1} \Delta Y_{t-1} + \sigma_t z_t$. We fill the initial values of the generated systems with draws from the standard normal distribution, and use them as initial values for the bootstrap replications of the systems as well (following Cavaliere et al. (2010a)).

For our simulations, we follow the latest literature standards and use $10,000$ replications to examine the $T = 100$ realisations case. The reasoning is that large-sample results of the involved tests include some sort of convergence of the tests' power towards unity, therefore only small-sample cases are likely to pose a challenge regarding performance. In order to catch any phenomena occurring as both the rank and dimension of the system change, we examine the cases of $p = 2, 3$ and 4 variables and all the corresponding possible

nulls for each case (i.e., $r = 0, 1$ for $p = 2$, $r = 0, 1, 2$ for $p = 3$ and so on). We measure the power of the tests as the probability of correctly rejecting or accepting a null hypothesis and the size as the probability of incorrectly rejecting a null hypothesis. The probability of incorrectly accepting a null is trivial to retrieve. Therefore, we have

$$\text{power} = \frac{\#\ \text{of correct rank estimations}}{10000} \tag{12}$$

$$\text{size} = \frac{\#\ \text{of over-rejections}}{10000} \tag{13}$$

Our simulations method for evaluating the performance of the tests for rank r and innovations p can be summed up with the following algorithm

Steps
While *successful replications* $\leq 10,000$
1. Generate innovations $\{\epsilon_t\}_{t=1}^{100}$, parameters A_r, R_r, D, Λ, initial values Y_1, Y_2.
2. Construct system $\{Y_t\}_{t=1}^{100}$ recursively following (11).
3. If $Y_t \sim I(1)$
a. Run test (LR/WB) sequentially for $r = 0, \ldots, p-1$ and save estimated rank (number of rejection of nulls).
b. If results available (e.g. not NULL due to explosiveness/seasonality).
i. Save estimated rank.
ii. *successful replications* $\leftarrow$ *successful replications* + 1
Power = correct rank estimations/10,000 , Size = rank over-estimations/10,000.

Table 8: Simulations methodology algorithm

For the bootstrap simulations, we use a 5% significance level and 500 bootstrap replications. Also, as discussed in section 2, we neglect deterministic terms such as constants or trends in order to ensure the feasibility of our simulations.

Unfortunately, despite applying restrictions to achieve computational feasibility, most of the systems generated in order to obtain $10,000$ replications (probably somewhere close to $50,000$) are discarded due to stationarity or explosiveness. To establish whether a system is suitable to apply our tests, we run the augmented Dickey-Fuller test and require that the statistic lies between zero and the critical value. For assessing whether a system is proper for applying bootstrap implementations of Johansen's test, we use the coefficients estimates of Johansen's test to formulate the coefficient matrix of the corresponding VECM's

companion form. We comply with Swensen (2006) and Cavaliere et al. (2010a,b) and require that the eigenvalues of the companion coefficient matrix either lie outside the unit circle or are equal to one, therefore there is no explosiveness or seasonality in the system.

4.5 Computational Details

All simulations are carried out with the R programming language, which combines an easy syntax, optimized procedures for statistical analysis (including vectorized operations, recursions, applications of functions etc.) and indispensable flexibility regarding functional programming. The code for simulation procedures was greatly enhanced and optimized with the invaluable help of Dr. Georgios Papapanagiotou, whose assistance has been critical in carrying out the present work. For Johansen's test, the code was produced by the author in accord with the extensive review of the test and its procedures given in Johansen (1995) and is, in part, inspired by the code of the *ca.jo* function of the *urca* time series analysis R package (see Pfaff (2008)). The code for the IID and wild bootstrap cointegration tests is also produced by the author, based on Swensen (2006), Cavaliere et al. (2010a), Cavaliere et al. (2010b), Cavaliere et al. (2012), Cavaliere et al. (2014) and Cavaliere et al. (2015). The bootstrap tests code has been enriched with elements of the code used in Cavaliere et al. (2015), kindly provided by Prof. Giuseppe Cavaliere.

At the culmination (and hopefully successful defence of) the present thesis, the author aspires to combine and refine the tools developed for this work, as well as some past independent ventures on cointegration testing, and produce an R package to complement the existing tools available in R for cointegration analysis.

5 Results

The following discussion summarises the outcomes of our simulations for the performance of Johansen's likelihood ratio test (LR) as well as its wild bootstrap (WB) implementation (Cavaliere et al. (2010a,b)). As stated above, we move from $p = 2$ to $p = 4$ catching all the in-between cases of cointegrating ranks, starting from $r = 0$ up to $r = p - 1$, hoping to reveal any trends or patterns that may emerge from this "continuous" assessment. We consider the case of 100 realisations and use 500 replications for the WB assessment.

Before examining the results, let us first discuss how we can properly evaluate them. As is the standard procedure in statistical hypothesis testing, the critical values for Johansen's test are chosen so that the test's size (probability of committing type I error) is equal to the chosen significance level (in this study 0.05). If the empirical size of the test exceeds 0.05, we can roughly declare that the test tends to over-reject the null hypothesis and therefore is less likely to correctly assess the number of long-run relations with our desired level of accuracy. For the power of the test, no such threshold past or below which statistical inference is compromised exists (since power is not directly connected to the statistical decision problem), therefore we simply desire that it is as high as possible. The following tables hold the results of our Monte Carlo simulations for assessing the power and size of Johansen's test, in accord to the methodology discussed in section 4.

5.1 Simulation Results

The results of table 9 hint towards some interesting conclusions regarding the performance of Johansen's likelihood ratio test. First, notice that, unlike the majority of the literature, our simulations have very small size distortions, with only the $p = 4$, $r = 0$ case exceeding the 5% significance level, where statistical decision making is compromised. On the other hand, there seems a tendency for type II errors to occur. Therefore, for the class of systems discussed here, Johansen's trace test has the tendency to sub-reject the null of the actual cointegrating rank.

The author speculates that this may be caused by the magnitude of the eigen-decomposition components of the memory coefficient of the VECM (Γ), since simulations with models of "weaker" memory (values that compose Γ drawn from smaller symmetric sub-intervals of $[-1, 1]$) yield much higher size distortions (results not reported here)[3]. The random choice of parameters, as well as the overall specification of the model are likely the cause for any inconsitency with the results of Lee and Tse (1996), Haug (1996) and others.

[3]Smaller memory coefficients also appear to have an effect on the performance of the wild bootstrap test (results not reported here).

Likelihood Ratio test

	Power				Size			
p	$r=0$	$r=1$	$r=2$	$r=3$	$r=0$	$r=1$	$r=2$	$r=3$
GARCH(1,1) Variance								
2	0.9775	0.8856	NA	NA	0.0225	0.0005	NA	NA
3	0.9512	0.8517	0.7410	NA	0.0488	0.0113	0.0014	NA
4	0.8908	0.8395	0.7043	0.6286	0.1092	0.0285	0.0115	0.0028
EGARCH(1,1) Variance								
2	0.9992	0.9805	NA	NA	0.0008	0.0000	NA	NA
3	0.9981	0.9834	0.9331	NA	0.0019	0.0003	0.0001	NA
4	0.9979	0.9868	0.9579	0.8944	0.0021	0.0011	0.0005	0.0002
Stochastic Volatility								
2	0.9997	0.9740	NA	NA	0.0003	0.0000	NA	NA
3	0.9995	0.9781	0.9204	NA	0.0005	0.0003	0.0000	NA
4	0.9991	0.9793	0.9456	0.8791	0.0009	0.0006	0.0004	0.0000
Outliers (3)								
2	0.9970	0.9958	NA	NA	0.0030	0.0000	NA	NA
3	0.9922	0.9971	0.9806	NA	0.0078	0.0021	0.0000	NA
4	0.9806	0.9964	0.9920	0.9694	0.0194	0.0025	0.0013	0.0001
Regime Shifts (1)								
2	0.9991	0.9968	NA	NA	0.0009	0.0000	NA	NA
3	0.9978	0.9976	0.9832	NA	0.0022	0.0011	0.0002	NA
4	0.9965	0.9974	0.9927	0.9671	0.0035	0.0022	0.0013	0.0007

Table 9: LR test performance, p variables, coint. rank r, 100 realisations

Notice that for cases of both high rank and a large number of variables ($r = 3,\ p = 4$), the power of the LR test decreases significantly. Also, the size of the test appears to increase with the number of variables but decrease with the number of long-run relations. The consistent presence of this phenomenon in our results leads to the conclusion that, at least for the class of models considered here, size will be higher

in larger cointegrated systems, and that the smaller the difference between cointegrating rank and dimesnion, the smaller the size of Johansen's LR test. It is worth commenting that the size of the test reaches 0.0000 on multiple occasions. This means that, for these cases, the test has not once in $10,000$ replications over-rejected the null hypothesis.

The resilience of the LR test's size against a variety of cointegrating ranks and dimensions does justify its wide acceptance and frequent use in applied econometrics, as, evidently, it is quite unlikely for the statistical decision process to fail. Of course, we should keep in mind that these results refer to a specific class of models, and the presence of additional factors such as an intercept, as well as the alteration of the intervals we use to draw values for our parameters could possibly lead to different results regarding power and size.

Let us now move to table 10, which holds the performance evaluation of the wild bootstrap test[4]. Here, we have no instance of the size ever exceeding the nominal significance level. Also, although not shown here, p-values produced by the WB test are "strict", in the sense that, for most cases, they appeared to not exceed to 0 unless the null was to be accepted. Also, neither the conclusion that size decreases with the number of long-run relations nor the result that it increases with the number of variables seem to hold consistently. Notice that the empirical size of the WB test is overall higher than that of the LR test, however sub-rejections of the null hypotheses (and therefore type II errors) appear to persist, and there are much fewer instances where size is estimated to be 0.0000.

Tendency for type II errors does contradict the literature on WB cointegration testing. However, this might also be attributed to the values of the Γ coefficient, as, for example, in Cavaliere et al. (2012), the lagged term coefficient matrix is almost diagonal and filled with small values. Simulations (run by the author, results not reported here) have shown that decreasing the magnitude of the coefficients of Γ shifts the error from type II to type I (over-sizing), where the possibility of compromising of statistical decision making is higher. Nevertheless, additional simulations ought to be carried out to determine whether this phenomenon occurs consistently or is simply circumstantial.

[4]Which was by far the hardest to run and complete.

Wild Bootstrap test

	Power				Size			
p	$r=0$	$r=1$	$r=2$	$r=3$	$r=0$	$r=1$	$r=2$	$r=3$
GARCH(1,1) Variance								
2	0.9904	0.9888	NA	NA	0.0096	0.0000	NA	NA
3	0.9924	0.9817	0.9512	NA	0.0076	0.0061	0.0010	NA
4	0.9869	0.9860	0.9544	0.9193	0.0131	0.0063	0.0101	0.0045
EGARCH(1,1) Variance								
2	0.9907	0.9930	NA	NA	0.0093	0.0000	NA	NA
3	0.9896	0.9854	0.9598	NA	0.0104	0.0074	0.0007	NA
4	0.9885	0.9872	0.9681	0.9327	0.0115	0.0059	0.0075	0.0031
Stochastic Volatility								
2	0.9947	0.9895	NA	NA	0.0053	0.0002	NA	NA
3	0.9934	0.9852	0.9586	NA	0.0066	0.0044	0.0003	NA
4	0.9954	0.9885	0.9564	0.9231	0.0046	0.0028	0.0063	0.0036
Outliers (3)								
2	0.9867	0.9988	NA	NA	0.0133	0.0002	NA	NA
3	0.9918	0.9896	0.9885	NA	0.0082	0.0096	0.0003	NA
4	0.9700	0.9823	0.9828	0.9794	0.0300	0.0176	0.0130	0.0013
Regime Shifts (1)								
2	0.9895	0.9969	NA	NA	0.0105	0.0010	NA	NA
3	0.9892	0.9856	0.9871	NA	0.0108	0.0132	0.0020	NA
4	0.9846	0.9826	0.9776	0.9735	0.0154	0.0173	0.0181	0.0061

Table 10: WB test performance, p variables, coint. rank r, 100 realisations

The results depicted in the following figures hold special interest. First, holding the number of variables constant at $p = 4$ and examining the path performance follows as r increases allows us to capture the effect of increasing number of cointegrating relations on the power of the likelihood ratio and wild bootstrap tests (figure 1). In that front, the WB test does seem to outperform the LR in all of the cases where the variance of innovations is

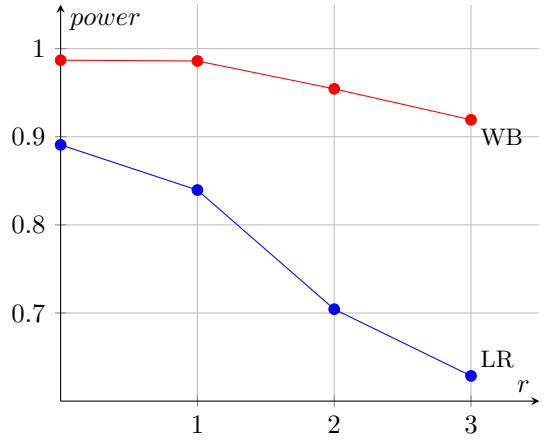
time dependent. The GARCH(1,1) innovations exhibit the largest performance differential, with WB consistently and significantly outperforming LR, while EGARCH(1,1) innovations and stochastic volatility seem to follow a similar path, where WB outperforms LR for all cointegrating ranks except for the $r = 0$ case. For the rest of the cases, applying the wild bootstrap test instead of the LR test does not seem to yield any significant benefit. In fact, for the cases of outliers and regime shifts, the LR test seems to have an overall better performance than the WB test. Also, both tests' power appears to monotonically decrease as the number of long-run relations increases. The only exceptions are LR - regime shift case and both WB and LR outlier cases. Strangely enough, the power of the WB test seems to initially increase as the cointegrating rank increases and only starts dropping at $r = 3$.

The difference in performance due to outliers in the variance may be caused by the lack of bias correction against specific observations, which is tied to the structure of the wild bootstrap. The author speculates that the IID and/or the block resampling schemes are likely to yield better results in such DGPs.

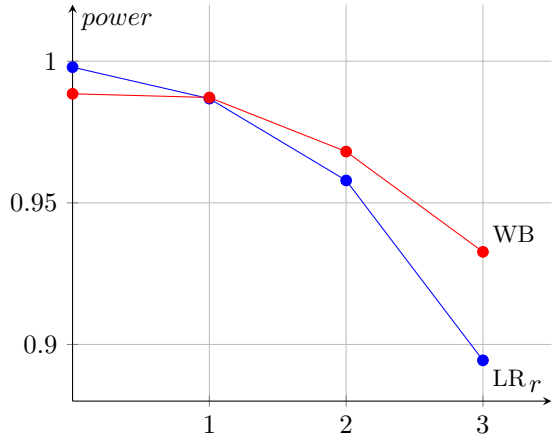
The second result emerges from holding the number of long-run relations constant and examining the power of the tests as the dimension of the model increases. Graphs in figure 2 show how power changes with respect to the number of variables when there are no cointegrating relations and when there exists one cointegrating relation[5]. First, notice that power distortions due to dimension never actually drop below 95%, therefore the validity of the statistical decision rule is not threatened. A tendency for power to decrease with dimension is present, but not consistent, as, for example, the power of the LR test seems to increase with dimension under $r = 1$. Also, in some instances, power seems to increase and then decrease, or the opposite. Notice that the LR test strictly outperforms the WB test in the case of a single regime shift for both $r = 0$ and $r = 1$ and in the EGARCH(1,1) case for $r = 0$. The only case where WB is strictly dominant, both with respect to rank and dimension, is the GARCH(1.1) case, despite our choice of GARCH coefficients summing up close to unity.

The results of our simulations indicate that, at least for the class of systems considered here, the WB test can mitigate the presence of conditional heteroskedasticity, mainly in its GARCH form, and also alleviate, to some extent, power loss due to other time-dependent specifications (EGARCH, stochastic volatility), but does not significantly improve inference in the presence of regime shifts or outliers. Examination with more regime shifts and outliers of varying magnitude could possibly result in a better understanding of both tests' behaviour under such instances.

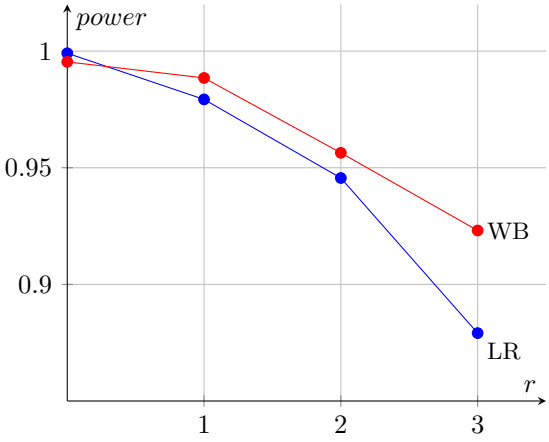
[5]We avoid the rest since no patterns can emerge from one or two datapoints.



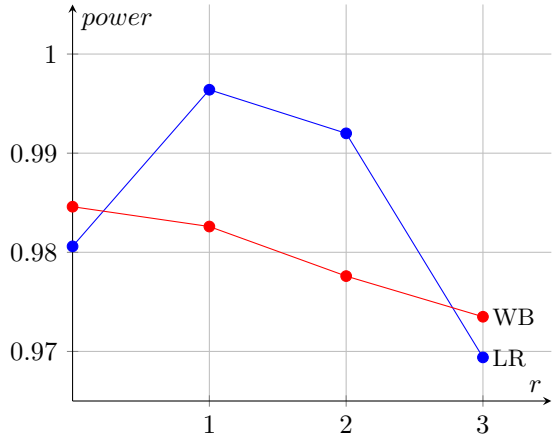
(a) GARCH(1,1) innovations



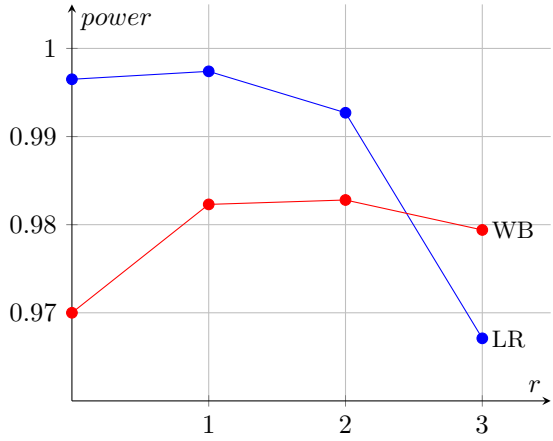
(b) EGARCH(1,1) innovations



(c) Stochastic Volatility

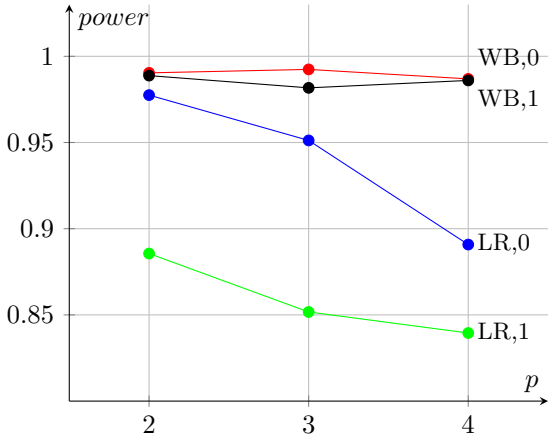


(d) Regime Shifts (1)

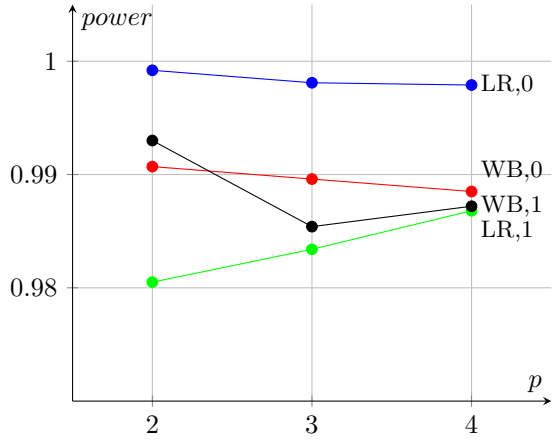


(e) Outliers (3)

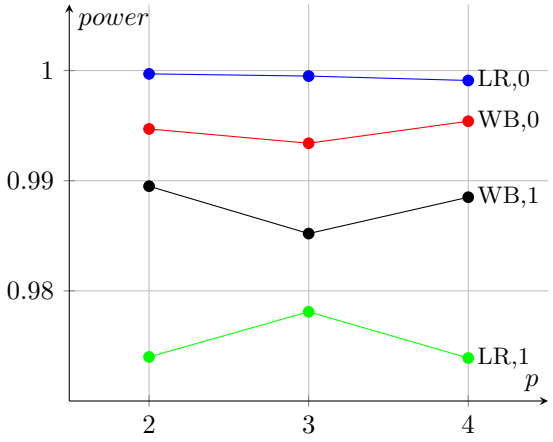
Figure 1: Power of likelihood ratio and wild bootstrap tests, $p = 4$, 100 realisations



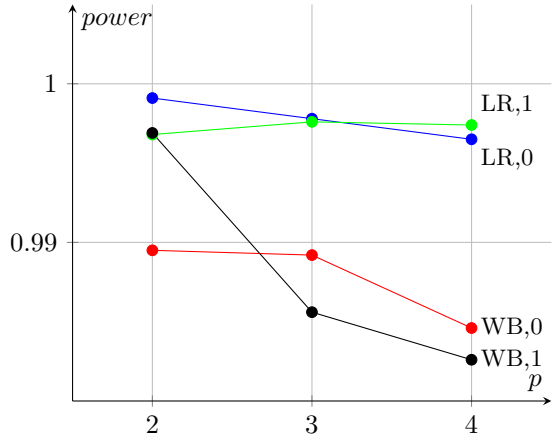
(a) GARCH(1,1) innovations



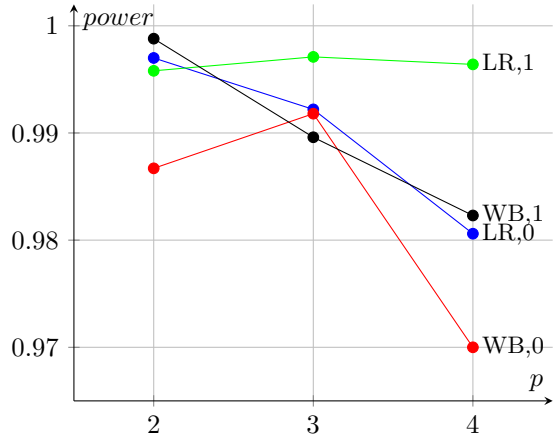
(b) EGARCH(1,1) innovations



(c) Stochastic Volatility



(d) Regime Shifts (1)



(e) Outliers (3)

Figure 2: Power of likelihood ratio and wild bootstrap tests, $r = 0$, $r = 1$, 100 realisations

5.2 Empirical Application

Before moving to conclusions, it is worth applying both tests to a real world system that is likely to exhibit some of the traits our discussion revolved around. We use daily observations of futures prices for crude oil and natural gas[6] for the year 2023. The ARCH test strongly indicates the presence of conditional heteroskedasticity, making both series promising candidates for the empirical exhibition of the difference in results we identified throughout this section. Also, both series are unit roots, which ensures all the nice large-sample results for $I(1)$ systems. The following table contains the LR and WB test results for the nulls $r = 0$ and $r = 1$.

H_0	WB p-values	LR trace stat	LR c.v (10%,	5%,	1%)
$r = 0$	0.020	25.2544	15.66	17.95	23.52
$r \leq 1$	0.698	4.3537	6.50	8.18	11.65

Table 11: LR and WB tests for crude oil and natural gas prices, 1/1/2023-1/1/2024

Evidently, at 5% statistical significance, both tests reject the null of no long-run relations and do not reject the null of at most one long-run relation. However, with a more "strict" approach of 1% statistical significance, the WB test does not actually reject the null of $r = 0$. Unfortunately, knowledge of the true cointegrating rank of the system is impossible. Nevertheless, this application demonstrates the possibility of the two tests leading to different inferences, which, in more intense datasets (e.g. cryptocurrencies, stocks etc.), could lead to significantly different estimates of the number of long-run relations.

[6]Traded at the New York Mercantile from 1/1/2023 to 1/1/2024, data from Yahoo finance.

6 Conclusions

6.1 Summary and Discussion

In this study, we attempted to evaluate the performance of the wild bootstrap implementation of Johansen's test of Cavaliere et al. (2010a,b) and compare it to the performance of Johansen's test. results from tables 9 and 10 show that the WB test is largely robust against conditional heteroskedasticity and stochastic volatility (at least for the specifications considered here) and may significantly out-perform the LR test, but also that the classical LR test outperforms the WB test in the presence of outliers and regime shifts. These results hold for the class of processes we have so far considered. Also, most failed estimations seem to be type II errors, since both tests exhibit an intense tendency to underestimate the number of long-run relations in the model. This may be the result of the chosen sets parameters are randomly drawn from. Overall, we can claim that the use of the wild bootstrap cointegration test is likely to yield more accurate results than the Johansen's likelihood ratio test when the series exhibit heteroskedasticity, especially as the number of variables grows, where the presence of more than one long-run relations is a possibility.

6.2 Directions for Further Study

A large number of cases that concern realistic models have not been discussed here. Misspecification with respect to lag length or the deterministic terms could have a significant effect on the performance of the WB test. Furthermore, especially in highly volatile series, one could expect the system to exhibit a Markov switching structure in either the deterministic part and/or the variance of the model. Of course, acknowledging all these factors in a single study would be nearly impossible, both in terms of computational feasibility and in terms of DGP generality. However, investigating these cases is likely to yield important results and let us map the suitability of each cointegration testing procedure under each of these traits.

There is also room for extensions in the innovations processes considered. The author has produced (but not yet run extensively) code for innovations with a dynamic conditional correlation structure (DCC) which would be useful for examining models connected not only equation-wise but also through correlation. In fact, a DCC approach with a specification similar to that of Maki (2013) would hold special interest for applications in finance. Multivariate GARCH structures, Markov-switching variance, as well as the incorporation of any number of regime shifts and outliers are also worth examining. Finally, combinations of these traits in DGPs' innovations would get us as close as one can be to actual, real-world systems. Such specifications would probably challenge our testing procedures (as well as the capacity of our computers' processors).

Given that our discussion revolves around bootstrap methods, it would be interesting and useful to investigate the performance of bootstrap cointegration tests under problematic sampling, which is frequently the case with macroeconomic data. Frequency reduction, impartial sampling and missing values are some pretty realistic instances in time series analysis, and are certainly worth exploring to evaluate the robustness of cointegration tests against improper samples, which constitute a sort of "natural enemy" for the bootstrap method. Additionally, since, under the general reduced rank specification of Johansen's model, cointegrating rank is essentially a model selection problem, it would be worth trying all of the above issues and factors of distortion against inference based on information criteria. Some important steps are made in Cheng and Phillips (2009), Cheng and Phillips (2012) and Cavaliere et al. (2015), however there is still room for testing with simulations against a variety of DGPs.

Another, relatively unexplored case where running simulations may yield fruitful results is that of changing cointegrating ranks throughout the timespan of the model. This form of structural breaks would possibly correspond to changes in preferences, changes in demographics, policy changes or intense market shocks, therefore appropriate inerene as well as robustness against temporal changes in the cointegrating rank are definitely worth exploring. Additionally, situations where not all variables are of the same order of integration should be considered since, as Johansen himself notes in Johansen (1995), econometric techniques should not govern what questions we ask or enforce us to re-shape our questions to fit their requirements.

References

Asai, M., McAleer, M., and Yu, J. (2006). Multivariate stochastic volatility: a review. *Econometric Reviews*, 25(2-3):145–175.

Bampinas, G., Ladopoulos, K., and Panagiotidis, T. (2018). A note on the estimated garch coefficients from the s&p1500 universe. *Applied Economics*, 50(34-35):3647–3653.

Bauwens, L., Deprins, D., Vandeuren, J.-P., et al. (1997). *Modelling interest rates with a cointegrated VAR-GARCH model*. Citeseer.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Boswijk, H. P., Cavaliere, G., Rahbek, A., and Taylor, A. M. R. (2016). Inference on co-integration parameters in heteroskedastic vector autoregressions. *Journal of Econometrics*, 192(1):64–85.

Boswijk, H. P., Jansson, M., and Nielsen, M. Ø. (2015). Improved likelihood ratio tests for cointegration rank in the var model. *Journal of Econometrics*, 184(1):97–110.

Breitung, J. (2002). Nonparametric tests for unit roots and cointegration. *Journal of econometrics*, 108(2):343–363.

Caldeira, J. and Moura, G. V. (2013). Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy. *Available at SSRN 2196391*.

Cavaliere, G., Angelis, L. D., Rahbek, A., and Robert Taylor, A. (2015). A comparison of sequential and information-based methods for determining the co-integration rank in heteroskedastic var models. *Oxford Bulletin of Economics and Statistics*, 77(1):106–128.

Cavaliere, G., De Angelis, L., Rahbek, A., and Taylor, A. M. R. (2016). Determining the Cointegration Rank in Heteroskedastic VAR Models of Unknown Order. *Econometric Theory*, 34(2):349–382.

Cavaliere, G., Nielsen, H. B., and Rahbek, A. (2020). An introduction to bootstrap theory in time series econometrics. *Available at SSRN 3589144*.

Cavaliere, G., Rahbek, A., and Robert Taylor, A. (2014). Bootstrap determination of the co-integration rank in heteroskedastic var models. *Econometric Reviews*, 33(5-6):606–650.

Cavaliere, G., Rahbek, A., and Taylor, A. M. R. (2010a). Testing for co-integration in vector autoregressions with non-stationary volatility. *Journal of Econometrics*, 158(1):7–24.

Cavaliere, G., Rahbek, A., and Taylor, A. R. (2010b). Cointegration rank testing under conditional heteroskedasticity. *Econometric Theory*, 26(6):1719–1760.

Cavaliere, G., Rahbek, A., and Taylor, A. R. (2012). Bootstrap determination of the co-integration rank in vector autoregressive models. *Econometrica*, 80(4):1721–1740.

Cavaliere, G., Rahbek, A., and Taylor, R. (2010c). Determination of the number of common stochastic trends under conditional heteroskedasticity. *Estudios de Economía Aplicada*, 28(3):519–551.

Cavaliere, G. and Taylor, R. (2006). Testing the null of co-integration in the presence of variance breaks. *Journal of Time Series Analysis*, 27(4):613–636.

Cheng, X. and Phillips, P. C. (2009). Semiparametric cointegrating rank selection. *The Econometrics Journal*, 12:S83–S104.

Cheng, X. and Phillips, P. C. (2012). Cointegrating rank selection in models with time-varying variance. *Journal of Econometrics*, 169(2):155–165.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431.

Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics: Methodology and distribution*, pages 569–593. Springer.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007.

Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276.

Fleming, J. and Kirby, C. (2003). A closer look at the relation between garch and stochastic autoregressive volatility. *Journal of financial econometrics*, 1(3):365–419.

Giersbergen, N. P. v. (1996). Bootstrapping the trace statistic in var models: Monte carlo results and applications. *Oxford Bulletin of Economics and Statistics*, 58(2):391–408.

Granger, C. W. (1981). Some properties of time series data and their use in econometric model specification. *Journal of econometrics*, 16(1):121–130.

Granger, C. W. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the s&p 500 absolute stock returns. *Journal of empirical finance*, 11(3):399–421.

Hansen, B. E. (1992). Heteroskedastic cointegration. *Journal of Econometrics*, 54(1-3):139–158.

Harris, R. I. and Judge, G. (1998). Small sample testing for cointegration using the bootstrap approach. *Economics Letters*, 58(1):31–37.

Haug, A. A. (1996). Tests for cointegration a Monte Carlo comparison. *Journal of Econometrics*, 71(1-2):89–115.

Haug, A. A., MacKinnon, J. G., and Michelis, L. (2000). European monetary union: a cointegration analysis. *Journal of International Money and Finance*, 19(3):419–432.

Höglund, R. and Östermark, R. (2003). Size and power of some cointegration tests under structural breaks and heteroskedastic noise. *Statistical Papers*, 44:1–22.

Horowitz, J. L. (2001). The bootstrap. In *Handbook of econometrics*, volume 5, pages 3159–3228. Elsevier.

Hubrich, K., Lütkepohl, H., and Saikkonen, P. (2001). A review of systems cointegration tests. *Econometric Reviews*, 20(3):247–318.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, 12(2-3):231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica: journal of the Econometric Society*, pages 1551–1580.

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

Jusélius, K. (2006). *The cointegrated VAR model : methodology and applications*.

Krolzig, H.-M. (1997). *Markov-Switching Vector Autoregressions: modelling, statistical inference, and application to business cycle analysis*.

Kurita, T. (2013). Exploring the impact of multivariate garch innovations on hypothesis testing for cointegrating vectors. *Communications in Statistics-Simulation and Computation*, 42(8):1785–1800.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis 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(1-3):159–178.

Lanne, M. and Saikkonen, P. (2005). Non-linear garch models for highly persistent volatility. *The Econometrics Journal*, 8(2):251–276.

Lee, T.-H. and Tse, Y. (1996). Cointegration tests with conditional heteroskedasticity. *Journal of Econometrics*, 73(2):401–410.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

MacKinnon, J. G., Haug, A. A., and Michelis, L. (1999). Numerical distribution functions of likelihood ratio tests for cointegration. *Journal of applied Econometrics*, 14(5):563–577.

Maddala, G. S. and Kim, I.-M. (1998). *Unit Roots, Cointegration, and Structural Change*. Cambridge University Press.

Maki, D. (2013). The influence of heteroskedastic variances on cointegration tests: A comparison using monte carlo simulations. *Computational Statistics*, 28(1):179–198.

Malmsten, H. and Teräsvirta, T. (2010). Stylized facts of financial time series and three popular models of volatility. *European Journal of pure and applied mathematics*, 3(3):443–477.

Mantalos, P. and Shukur, G. (1999). Testing for cointegrating relations-a bootstrap approach.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the econometric society*, pages 347–370.

Nielsen, H. B. (2004). Cointegration analysis in the presence of outliers. *The Econometrics Journal*, 7(1):249–271.

Nyblom, J. and Harvey, A. (2000). Tests of common stochastic trends. *Econometric theory*, 16(2):176–199.

Osterwald-Lenum, M. (1992). A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics 1. *Oxford Bulletin of Economics and statistics*, 54(3):461–472.

Otero, J., Panagiotidis, T., and Papapanagiotou, G. (2022). Multivariate cointegration and temporal aggregation: some further simulation results. *Computational Economics*, pages 1–12.

Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics*. Oxford University Press, USA.

Peter Boswijk, H., Cavaliere, G., De Angelis, L., and Taylor, A. R. (2023). Adaptive information-based methods for determining the co-integration rank in heteroskedastic var models. *Econometric Reviews*, 42(9-10):725–757.

Pfaff, B. (2008). *Analysis of Integrated and Cointegrated Time Series with R*. Springer.

Phillips, P. C. and Ouliaris, S. (1988). Testing for cointegration using principal components methods. *Journal of Economic Dynamics and Control*, 12(2-3):205–230.

Phillips, P. C. and Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica: journal of the Econometric Society*, pages 165–193.

Shin, Y. (1994). A residual-based test of the null of cointegration against the alternative of no cointegration. *Econometric theory*, 10(1):91–115.

Silvapulle, P. and Podivinsky, J. M. (2000). The effect of non-normal disturbances and conditional heteroskedasticity on multiple cointegration tests. *Journal of Statistical Computation and Simulation*, 65(1-4):173–189.

Stock, J. H. and Watson, M. W. (1988). Testing for common trends. *Journal of the American statistical Association*, 83(404):1097–1107.

Swensen, A. R. (2006). Bootstrap algorithms for testing and determining the cointegration rank in var models 1. *Econometrica*, 74(6):1699–1714.

Tanaka, K. (2017). *Time series analysis: nonstationary and noninvertible distribution theory*. John Wiley & Sons.

Vücel, M. K. and Guo, S. (1994). Fuel taxes and cointegration of energy prices. *Contemporary Economic Policy*, 12(3):33–41.