

University of Duisburg-Essen  
Faculty of Business Administration and  
Economics  
Chair of Econometrics



# P-Approximation

Seminar in Econometrics

Term Paper

Submitted to the Faculty of  
Business Administration and Economics  
at the  
University of Duisburg-Essen

from:

Jens Klenke and Janine Langerbein

---

Reviewer: Christoph Hanck

Deadline: Jan. 17th 2020

---

Name:	Jens Klenke	Janine Langerbein
Matriculation Number:	3071594	307
E-Mail:	jens.klenke@stud.uni-due.de	janine.langerbein@stud.uni-due.de
Study Path:	M.Sc. Economics	M.Sc. Economics
Semester:	5 <sup>th</sup>	5 <sup>th</sup>
Graduation (est.):	Winter Term 2020	Winter Term 2020

# Contents

List of Figures	II
List of Tables	II
List of Abbreviations	II
1 Introduction	1
2 Bayer Hanck Test	1
3 Simulation	3
4 Models	4
5 Package	4
References	I
Software-References	IV
A Appendices	V

**List of Figures**

**List of Tables**

**List of Abbreviations**

# 1 Introduction

Meta tests have been shown to be a powerful tool when testing for the null of non-cointegration. The distribution of their test statistic, however, is mostly not available in closed form. This might pose difficulties when implementing the meta tests in econometric software packages, as one has to include tables of critical values and p-values for each combination of the underlying tests. Software package size limitations are therefore quickly exceeded.

In this paper we propose supervised Machine Learning Algorithms to approximate the p-values of the meta test by Bayer and Hanck (2012) which tests for the null of non-cointegration. This approach might reduce the size of associated software packages considerably. The algorithms are trained on simulated data for various specifications of the aforementioned test.

Ergebnis der Models (1-2 Sätze)

Inhalt Paper

## 2 Bayer Hanck Test

The choice as to which of the available cointegration tests to use is a recurrent issue in econometric time series analysis. Bayer and Hanck (2012) propose powerful meta tests which provide unambiguous test decisions. They combine several residual- and system-based tests in the manner of Fisher's (1932) Chi-squared test.

Bayer and Hanck build their work on results from Pesavento (2004), who defines the underlying model as  $z'_t = [x'_t, y_t]$ .  $x_t$ , an  $n_1 \times 1$  vector, describes the regressor dynamics, while  $y_t$  is a scalar which defines the cointegrating relation. They can be written as

$$\Delta x_t = \tau_1 + v_{1t}, \quad (2.1)$$

$$y_t = (\mu_2 - \gamma' \mu_1) + (\tau_2 - \gamma' \tau_1)t + \gamma' x_t + u_t, \quad (2.2)$$

$$u_t = \rho u_{t-1} + v_{2t}. \quad (2.3)$$

$\mu_1$ ,  $\mu_2$ ,  $\tau_1$  and  $\tau_2$  are the deterministic parts of the model. They are subject to the following restrictions: (i)  $\mu_2 - \gamma' \mu_1$  and  $\tau = 0$  which translates to no deterministic, (ii)  $\tau = 0$  which corresponds to a constant in the cointegrating vector, (iii)  $\tau_2 - \gamma' \tau_1 = 0$ , a constant plus trend.

$v_t = [v'_{1t} v_{2t}]'$  with  $\Omega$  the long-run covariance matrix of  $v_t$ . For derivation of  $v_t$  see Pesavento (2004). Pesavento shows that  $\{v_t\}$  satisfies an FCLT, i.e.  $T^{-1/2} \sum_{t=1}^{[T\cdot]} v_t \Rightarrow \Omega^{1/2} W(\cdot)$ . It is further assumed that the  $x_t$  are not cointegrated.

It clearly follows from (2.3) that  $z_t$  is cointegrated if  $\rho < 1$ . Hence the null hypothesis of no cointegration is  $H_0 : \rho = 1$ .

Furthermore, Pesavento introduces two other parameters. First,  $R^2$  measures the squared correlation of  $v_{1t}$  and  $v_{2t}$ . It can be interpreted as the influence of the right-hand side variables in (2.2). It ranks between zero and one. When there is no long-run correlation between those variables and the errors from the cointegration regression,  $R^2$  equals zero. Secondly, the number of lags is approximated by a finite number  $k$ .

### Assumptions (BH S. 84)?

Bayer and Hanck's (2012) meta test combines the test statistics of four stand-alone tests. Namely, these are the tests of Engle and Granger (1987), Johansen (1988), Boswijk (1994) and Banerjee, Dolado, and Mestre (1998). For the sake of brevity the detailed derivation of the underlying tests has been deliberately omitted here.

Engle and Granger (1987) propose a two-step procedure to test the null hypothesis of no cointegration against the alternative of at least one cointegrating vector. First, the long-run relationship between  $y_t$  and  $\mathbf{x}_t$  is estimated by least squares regression. The obtained residuals  $\hat{u}_t$  are then tested for a unit root. For this, Engle and Granger suggest the use of the  $t$ -statistic  $t_\gamma^{\text{ADF}}$  in the Augmented Dickey-Fuller (ADF) regression:

$$\Delta \hat{u}_t = \gamma \hat{u}_{t-1} + \sum_{i=1}^k \pi_i \Delta \hat{u}_{t-i} + \varepsilon_t. \quad (2.4)$$

The rejection of a unit root points to a cointegration relationship.

Johansen's (1988) maximum eigenvalue test is a system-based test that allows for several cointegration relationships. Take the vector error correction model (VECM)

$$\Delta \mathbf{z}_t = \Pi \mathbf{z}_{t-1} + \sum_{i=1}^k \Gamma_i \Delta \mathbf{z}_{t-i} + \mathbf{d}_t + \varepsilon_t. \quad (2.5)$$

blabla Johansen test statistic

Banerjee and Boswijk

To combine the results from the underlying tests Bayer and Hanck draw upon Fisher’s combined probability test (Fisher, 1932). It merges the tests using the formula

$$\tilde{\chi}_{\mathcal{I}}^2 := -2 \sum_{i \in \mathcal{I}} \ln(p_i). \quad (2.6)$$

Let  $t_i$  be the  $i^{th}$  test statistic. If test  $i$  rejects for large values, take  $\xi_i := t_i$ . If test  $i$  rejects for small values, take  $-\xi_i := t_i$ . With  $\Xi_i(x) := \Pr_{\mathcal{H}_0}(\xi_i \geq x)$  the p-value of the  $i^{th}$  test is  $p_i := \Xi_i(\xi_i)$ .

Fisher (1932) shows that under the assumption of independence the null distribution of  $\tilde{\chi}_{\mathcal{I}}^2$  follows a chi-squared distribution with  $2\mathcal{I}$  degrees of freedom. If this assumption is violated the null distribution is less evident. Here, the latter case occurs, as the  $\xi_i$  are not independent. The  $\tilde{\chi}_{\mathcal{I}}^2$ , however, have well-defined asymptotic null distributions  $F_{\mathcal{F}_{\mathcal{I}}}$ , as  $\tilde{\chi}_{\mathcal{I}}^2 \rightarrow_d \mathcal{F}_{\mathcal{I}}$  under  $\mathcal{H}_0$  if  $T \rightarrow \infty$ , with  $\mathcal{F}_{\mathcal{I}}$  some random variable. It is therefore feasible to simulate the joint null distribution of the  $\xi_i$  to obtain the distribution  $F_{\mathcal{F}_{\mathcal{I}}}$  of 2.6. The  $F_{\mathcal{F}_{\mathcal{I}}}$  depend on which and how many tests are combined. The distributions of the  $\xi_i$  depend on  $K - 1$  and the deterministic case.

### 3 Simulation

In this section, we describe the simulation of the null distribution of the Bayer Hanck meta test. The objective is to obtain data for training machine learning algorithms on approximating the p-values of the aforementioned test. In consideration of the different forms of the meta test we generated six data sets. These vary according to the specific combination of the underlying tests and also account for the above-mentioned restrictions on the deterministic parts of the model.

This simulation relies largely on previous work by Pesavento (2004). We consider  $R^2 \in \{0, 0.05, 0.1, \dots, 0.95\}$ ,  $k = 11$  and  $c = 0^1$  and set the number of repetitions to 1,000,000. **N? c vielleicht mal definieren**

To calculate the Bayer Hanck test statistic we first simulate the null distributions of the underlying test statistics. It can be shown that asymptotically these are non-standard but a function of standard Brownian motions. The latter is approximated by step functions using Gaussian random walk with

---

<sup>1</sup>Since we solely aim at simulating the distribution of the null of no cointegration we will not consider any further values of  $c$  here.

$N = 1000$  observations. Referenz Theorem? OU Prozess? Nochmal auf Transformation je nach case eingehen? p values durch cdf, Fisherstat + pvalues berechnen.

## 4 Models

## 5 Package

## References

- Banerjee, A., Dolado, J., & Mestre, R. (1998). Error-correction mechanism tests for cointegration in a single-equation framework. *Journal of Time Series Analysis*, 19(3), 267–283. Retrieved from <https://EconPapers.repec.org/RePEc:bla:jtsera:v:19:y:1998:i:3:p:267-283>
- Bayer, C., & Hanck, C. (2012). Combining non-cointegration tests. *Journal of Time Series Analysis*.
- Boswijk, H. P. (1994). Testing for an unstable root in conditional and structural error correction models. *Journal of Econometrics*, 63(1), 37–60. Retrieved from <https://EconPapers.repec.org/RePEc:eee:econom:v:63:y:1994:i:1:p:37-60>
- Engle, R., & Granger, C. W. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55, 251–276.
- Fisher, R. A. (1932). *Statistical methods for research workers*. Oliver, Boyd, Edinburgh, and London.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2), 231–254. doi:[https://doi.org/10.1016/0165-1889\(88\)90041-3](https://doi.org/10.1016/0165-1889(88)90041-3)
- Pesavento, E. (2004). Analytical evaluation of the power of tests for the absence of cointegration. *Journal of Econometrics*, 122(2), 349–384.



## Software-References

- Breiman, L., Cutler, A., Liaw, A., & Wiener, M. (2018). *Randomforest: Breiman and cutler's random forests for classification and regression*. R package version 4.6-14. Retrieved from <https://CRAN.R-project.org/package=randomForest>
- Croissant, Y., Millo, G., & Tappe, K. (2019). *Plm: Linear models for panel data*. R package version 2.1-0. Retrieved from <https://CRAN.R-project.org/package=plm>
- Friedman, J., Hastie, T., Tibshirani, R., Simon, N., Narasimhan, B., & Qian, J. (2019). *Glmnet: Lasso and elastic-net regularized generalized linear models*. R package version 2.0-18. Retrieved from <https://CRAN.R-project.org/package=glmnet>
- Greenwell, B., Boehmke, B., Cunningham, J., & Developers, G. (2019). *Gbm: Generalized boosted regression models*. R package version 2.1.5. Retrieved from <https://CRAN.R-project.org/package=gbm>
- Henry, L., & Wickham, H. (2019). *Purrr: Functional programming tools*. R package version 0.3.2. Retrieved from <https://CRAN.R-project.org/package=purrr>
- Hlavac, M. (2018). *Stargazer: Well-formatted regression and summary statistics tables*. R package version 5.2.2. Retrieved from <https://CRAN.R-project.org/package=stargazer>
- Izrailev, S. (2014). *Tictoc: Functions for timing r scripts, as well as implementations of stack and list structures*. R package version 1.0. Retrieved from <https://CRAN.R-project.org/package=tictoc>
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., ... Hunt, T. (2019). *Caret: Classification and regression training*. R package version 6.0-84. Retrieved from <https://CRAN.R-project.org/package=caret>
- Lumley, T., & Miller, A. (2017). *Leaps: Regression subset selection*. R package version 3.0. Retrieved from <https://CRAN.R-project.org/package=leaps>
- Mevik, B.-H., Wehrens, R., & Liland, K. H. (2019). *Pls: Partial least squares and principal component regression*. R package version 2.7-1. Retrieved from <https://CRAN.R-project.org/package=pls>

- Milborrow, S. (2019a). *Plotmo: Plot a model's residuals, response, and partial dependence plots*. R package version 3.5.5. Retrieved from <https://CRAN.R-project.org/package=plotmo>
- Milborrow, S. (2019b). *Rpart.plot: Plot 'rpart' models: An enhanced version of 'plot.rpart'*. R package version 3.0.7. Retrieved from <https://CRAN.R-project.org/package=rpart.plot>
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Ripley, B. (2019a). *Class: Functions for classification*. R package version 7.3-15. Retrieved from <https://CRAN.R-project.org/package=class>
- Ripley, B. (2019b). *Mass: Support functions and datasets for venables and ripley's mass*. R package version 7.3-51.4. Retrieved from <https://CRAN.R-project.org/package=MASS>
- Ripley, B. (2019c). *Tree: Classification and regression trees*. R package version 1.0-40. Retrieved from <https://CRAN.R-project.org/package=tree>
- RStudio Team. (2019). *Rstudio: Integrated development environment for r*. Version 1.2.1541. RStudio, Inc. Boston, MA. Retrieved from <http://www.rstudio.com/>
- Rushworth, A. (2019). *Inspectdf: Inspection, comparison and visualisation of data frames*. R package version 0.0.4. Retrieved from <https://CRAN.R-project.org/package=inspectdf>
- Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., & Despouy, P. (2019). *Plotly: Create interactive web graphics via 'plotly.js'*. R package version 4.9.0. Retrieved from <https://CRAN.R-project.org/package=plotly>
- Therneau, T., & Atkinson, B. (2019). *Rpart: Recursive partitioning and regression trees*. R package version 4.1-15. Retrieved from <https://CRAN.R-project.org/package=rpart>
- Ushey, K., Allaire, J., Wickham, H., & Ritchie, G. (2019). *Rstudioapi: Safely access the rstudio api*. R package version 0.10. Retrieved from <https://CRAN.R-project.org/package=rstudioapi>

- Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations*. R package version 1.4.0. Retrieved from <https://CRAN.R-project.org/package=stringr>
- Wickham, H., François, R., Henry, L., & Müller, K. (2019). *Dplyr: A grammar of data manipulation*. R package version 0.8.0.1. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- Wickham, H., & Henry, L. (2019). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. R package version 0.8.3. Retrieved from <https://CRAN.R-project.org/package=tidyr>
- Xie, Y. (2019). *Knitr: A general-purpose package for dynamic report generation in r*. R package version 1.23. Retrieved from <https://CRAN.R-project.org/package=knitr>

# A Appendices

Better sorting of the Appendix

### **Eidesstattliche Versicherung**

Ich versichere an Eides statt durch meine Unterschrift, dass ich die vorstehende Arbeit selbständig und ohne fremde Hilfe angefertigt und alle Stellen, die ich wörtlich oder annähernd wörtlich aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe. Die Arbeit hat in dieser oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen.

Essen, den \_\_\_\_\_

\_\_\_\_\_  
Jens Klenke and Janine Langerbein