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	3061371
E-Mail:	jens.klenke@stud.uni-due.de	janine.langerbein@stud.uni-due.de
Study Path:	M.Sc. Economics	M.Sc. Economics
Semester:	5th	5th
Graduation (est.):	Summer Term 2021	Summer Term 2021

Contents

List of Figures

List of Tables

List of Abbreviations

Lasso	Least Absolute Shrinkage and Selection Operator .	I
CDF	cumulative distribution function	4
RMSE	Root Mean Squared Error	4

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 the full null distribution 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 paper on previous work from Pesavento (2004), who defines the underlying model as $z_t' = [x_t', y_t]$, with x_t being an $n_1 \times 1$ vector and y_t a scalar, which displays the cointegration relation. They can be written as

$$\begin{aligned} \Delta x_t &= \tau_1 + v_{1t} && (2.1) \\ y_t &= (\mu_2 - \gamma' \mu_1) + (\tau_2 - \gamma' \tau_1)t + \gamma' x_t + u_t, && (2.2) \\ u_t &= \rho u_{t-1} + v_{2t}. && (2.3) \end{aligned}$$

Δx_t presents the regressor dynamics. μ_1, μ_2, τ_1 and τ_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 deterministics, (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 Ω 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 $H_0 : p = 1$. 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 considers the test statistics of up to four stand-alone tests. Namely, these are the tests of Engle and Granger (1987), Johansen (1988), Boswijk (1994) and Banerjee et al. (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_γ^{ADF} in the Augmented Dickey-Fuller (ADF) regression:

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

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)[1]

$$\Delta \mathbf{z}_t = \mathbf{\Pi} \mathbf{z}_{t-1} + \sum_{i=1}^{k} \mathbf{\Gamma}_p \Delta \mathbf{z}_{t-p} + \mathbf{d}_t + \varepsilon_t. \tag{2.5}$$

[1] Due to practical reasons we omit the derivation of the VECM which is presumed to be known.

We base this test on the test statistic $\lambda_{\max} = -T \ln(1 - \hat{\lambda}_t)$. π-Teil von BH?

The third and fourth test considered are error correction-based. Both estimate the equation

$$\Delta y_t = d_t + \pi'_{0x} \Delta x_t + \varphi_0 y_{t-1} + \varphi'_1 x_{t-1} + \sum_{p=1}^{P} (\pi'_{px} \Delta x_{t-p} + \pi_{py} \Delta y_{t-p}) \tag{2.6}$$

by ordinary least squares (OLS). Banerjee et al. (1998) then test the null of non-cointegration by applying a t-test on φ_0, i.e. $\mathcal{H}_0 : \varphi_0 = 0$?. Boswijk (1994) uses the Wald statistic for testing $\mathcal{H}_0 : (\varphi_0, \phi'_1)' = 0$.

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

$$\tilde{\chi}^2_{\mathcal{I}} := -2 \sum_{i \in \mathcal{I}} \ln{(p_i)}. \tag{2.7}$$

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 $\Xi_i(x) := \Pr_{\mathcal{H},}(\xi_i \geq x)$ 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 $\tilde{\chi}^2_{\mathcal{I}}$ 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 ξ_i are not independent. The $\tilde{\chi}^2_{\mathcal{I}}$, however, have well-defined asymptotic null distributions $F_{\mathcal{F}_\mathcal{I}}$, as $\tilde{\chi}^2_{\mathcal{I}} \to_d \mathcal{F}_\mathcal{I}$ under $\mathcal{H}_0$ if $T \to \infty$, with $\mathcal{F}_\mathcal{I}$ some random variable. It is therefore feasible to simulate the joint null distribution of the ξ_i to obtain the distribution $F_{\mathcal{F}_\mathcal{I}}$ of (2.7). The $F_{\mathcal{F}_\mathcal{I}}$ depend on which and how many tests are combined. The distributions of the ξ_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 generate six data sets. These vary according to the specific combinations of the underlying tests and also account for the above-mentioned restrictions on the deterministic parts of the model.

The following approach relies largely on previous work by Pesavento (2004). For calculating the Bayer Hanck test statistic we require the p-values of the underlying tests. For this, we simulate their null distributions. It can be shown that asymptotically these are functions of standard Brownian motions. Here, the latter are constructed by step functions using Gaussian random walk of size $N = 1000$. The number of repetitions is set to 1,000,000. Furthermore, we consider $\mathrm{R}^2 \in \{0, 0.05, 0.1, ..., 0.95\}$, the maximum number of lags $K = 11$ and $c = 0$[2] (c mal definieren).

From the mass of test statistics we build the cumulative distribution function (CDF) of each underlying test and calculate the respective p-values. These are inserted into (2.6) to eventually obtain the Bayer Hanck test statistics. Analogous to the previous approach, we deduce the associated null distribution and the p-values.

4 Models

We now use the generated data for training machine learning algorithms on predicting the approximated empirical CDF of the Bayer Hanck test. We work with the values of the test statistic and the number of lags k as predictors. As it is our objective to describe the null distribution with a less memory-intensive model we will only consider linear methods. For the same objective we compare the models according to their in-sample Root Mean Squared Error (RMSE). The threat of overfitting is thus of no particular relevance here. For this reason, and to reduce computation time, we use no cross-validation.

As the empirical CDF is typically known to be curved in an S-shape we skip the classic linear regression in favor of a more flexible model. We stay with least squares regression, but try various combinations of polynomial functions and interaction terms of the aforementioned regressors. The search for the best model is carried out via brute-force.

4.1 Data Pre-Processing

One approach for improving a model's predictive ability is the pre-processing of the training data. Some models, like linear regression, react sensitively to

[2] Since we solely aim at simulating the distribution of the null of no cointegration we will not consider any further values of c here.

certain characteristics of the predictor or response data. Those characteristics include, inter alia, distributional skewness and outliers and there exist several methods to lower their potentially bad impact on the model's performance.

Since we simulated our training data under the null of non-cointegration we expect the distribution of the test statistic to be rather right skewed. Plot also reveals it to have a long right tail. If we train our regression model on this raw data it can possibly have difficulties predicting from high values of the test statistic.

One of the aforementioned methods to deal with such issues are power transforms. One might decide freely which transformation to apply. Alternatively, there exist statistical methods to determine an appropriate transformation. A well-known family of transformations to un-skew data is the Box-Cox transformation (Box & Cox, 1964). They aim at transforming the data so that it closely resembles the normal distribution. The exact transformation depends on the parameter λ, whose optimal value can be empirically estimated:

$$y^{(\lambda)} = \begin{cases} \frac{y^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ \log(y), & \lambda = 0 \end{cases} \tag{4.1}$$

It is visible from (4.1) that Box and Cox (1964) developed these transformations for the dependent variable. Kuhn and Johnson (2013), however, report that it proves as effective for transforming individual regressors. We estimate lambda for the values of the test statistics of the Bayer-Hanck test and transform them according to (2.7). This forces their distribution into a more symmetric form.

Since the response variable consists of our p-values, which were simulated under the null hypothesis, it follows a uniform distribution and is already symmetric. A transformation would therefore not bring any apparent advantage. However, we still add a Box-Cox transformed and a logarithmised version of the response variable to see if it benefits the prediction.

We also include various variations of the actual categorical variable k. It is firstly decomposed into dummy variables and secondly recode as a numeric, so that various transformations can be performed.

4.2 Polynomial Regression

Due to the reasons given above we restrict ourselves to linear models. The empirical CDF, which we aim to predict, is known to have a curved shape. For this reason, a simple linear regression model is very unlikely to provide a satisfactory fit to the data. We are in need of a more flexible model to predict the response as accurately as possible.

Polynomial Regression extends the classic linear regression model by fitting a polynomial equation of arbitrary order to the data. A polynomial regression with n degrees thus takes the form

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + ... + \beta_n x_i^n + \varepsilon_i, \tag{4.2}$$

where ε_i is the error term. Quelle?

Here, we calculate orthogonal polynomials of the test statistic of the Bayer-Hanck Test, considering up to 15 degrees. We estimate the parameters with OLS. To potentially increase the predictive performance of our model we also add interaction terms and different transformations of the regressor k. Appendix lists all calculated models. Since there is no need to prevent overfitting we expect higher order polynomials to perform best, as they are highly flexible. These polynomials, however, tend to show a wiggly behaviour at the boundaries. This makes extrapolation beyond the limits of our simulated data a risky endeavour. We will address and fix this issue later on.

4.3 Lasso

As mentioned above our polynomial regression models are likely to perform best with higher order polynomials. With each added polynomial, however, we increase the complexity of our model and potentially add redundant regressors. Although, still, overfitting plays no major role here, we generally prefer sparser models in case of equal results. One way to deal with this is the use of variable selection methods. A well-known example of such methods is the Lasso.

The lasso estimate is defined as

$$\hat{\beta}^{\text{lasso}} = \arg\min_{\beta} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 \text{ s.t. } \sum_{j=1}^{p} |\beta_j| \leq t, \tag{4.3}$$

where the first term describes the residual sum of squares, subject to a term known as L1 penalty. In its Lagrangian form this can be rewritten as

$$\hat{\beta}^{\text{lasso}} = \arg\min_{\beta} \frac{1}{2} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{4.4}$$

λ is a tuning parameter which defines the degree of regularisation. The lasso penalty shrinks the coefficients and, for λ sufficiently large, can set them to zero. The value of λ is data dependent and is usually estimated with cross-validation. ausführlicher? Quelle?

We plan on fitting a LASSO model to polynomials of grade 15. We consider the same transformations and interaction terms as in earlier steps. We therefore fit a total of Anzahl models.

4.4 Other Regression Models

We also considered various other regression models. For different reasons they were not too suitable for our use case. Conventional non-linear methods, like Generalized Additive Models or Multivariate Adaptive Regression Splines, might have provided a decent prediction. However, the fitted models take up more memory space than the aforementioned linear methods. For the same reason refrain from using tree based methods. In addition, the latter tend to perform poorly with such a small amount of regressors. Given these limitations, we decided to stick solely with linear regression models.

5 Model Evaluation

We estimate all models for two different combinations of the underlying tests. Namely, these are a combination of the Engle-Granger and Johansen test (EJ) and a combination of all four underlying tests (all). Furthermore, we estimate one model per specification of the model deterministics. Altogether, this results in a total of six different models.

5.1 RMSE comparison

To measure the performance of our regression models we calculate their in-sample RMSE. This is an indication of how far the residuals of the models are from zero, with lower values preferable. We calculate the RMSE for predictions on the full distribution, as well as predictions on the lower tail ($p \leq 0.2$), as it is more important for the test decision of the Bayer-Hanck test. We also add a corrected version of the RMSE, cRMSE, where predictions are limited to [0, 1].

Table A3 lists all variations of the RMSE for the calculated polynomial regression models. It becomes apparent that a combination of higher order polynomials, dummy variables and interaction terms indeed achieves superior results compared to simpler models. For all variations of the RMSE the best models require a polynomial of minimum grade 12. That was to be expected, considering we are optimising an in-sample fit. The transformation of the response variable only seems to play a minor role in prediction accuracy. Interestingly, there are no major differences in model selection depending on the variation of the RMSE used. Table A2 lists the five best models for each case and test type.

For the above-mentioned reasons we choose the final models according to the cRMSE on the left tail of the distribution of the p-values. Grafik mit den 6 final models. It is apparent that the functional forms look very similar over all cases, mostly using the highest order polynomial available[3]. Furthermore, five out of six models use the Box-Cox transformed response variable.

5.2 Correction

6 Package

[3] We are well aware of the fact that this represents a corner solution. Since we are optimising the models on the in-sample RMSE, however, we could continue adding higher order polynomials forever to improve the fit. As we are already tweaking on the fifth decimal place we decided to not further pursue this procedure.

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. 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. https://EconPapers.repec.org/RePEc:eee:econom:v:63:y:1994:i:1:p:37-60

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*(2), 211–252. http://www.jstor.org/stable/2984418

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; London.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, *12*(2), 231–254. https://doi.org/https://doi.org/10.1016/0165-1889(88)90041-3

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer New York. https://books.google.de/books?id=xYRDAAAAQBAJ

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]. 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]. 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]. 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]. https://CRAN.R-project.org/package=gbm

Henry, L., & Wickham, H. (2019). *Purrr: Functional programming tools* [R package version 0.3.2]. https://CRAN.R-project.org/package=purrr

Hlavac, M. (2018). *Stargazer: Well-formatted regression and summary statistics tables* [R package version 5.2.2]. 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]. https://CRAN.R-project.org/package=tictoc

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., the R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., & Hunt., T. (2019). *Caret: Classification and regression training* [R package version 6.0-84]. https://CRAN.R-project.org/package=caret

Lumley, T., & Miller, A. (2017). *Leaps: Regression subset selection* [R package version 3.0]. 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]. 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]. 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]. 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. https://www.R-project.org/

Ripley, B. (2019a). *Class: Functions for classification* [R package version 7.3-15]. 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]. https://CRAN.R-project.org/package=MASS

Ripley, B. (2019c). *Tree: Classification and regression trees* [R package version 1.0-40]. https://CRAN.R-project.org/package=tree

RStudio Team. (2019). *Rstudio: Integrated development environment for r* [Version 1.2.1541]. RStudio, Inc. Boston, MA. http://www.rstudio.com/

Rushworth, A. (2019). *Inspectdf: Inspection, comparison and visualisation of data frames* [R package version 0.0.4]. 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]. https://CRAN.R-project.org/package=plotly

Therneau, T., & Atkinson, B. (2019). *Rpart: Recursive partitioning and regression trees* [R package version 4.1-15]. 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]. https://CRAN.R-project.org/package=rstudioapi

Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations* [R package version 1.4.0]. 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]. 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]. 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]. https://CRAN.R-project.org/package=knitr

A Appendices

Table A1 list the different functional forms of the polynomial regression we tested. In total we investigated 21 different forms and for each of these forms we investigated the polynomial in the range from 3 to 13. As equations with many polynomials are getting very long we will use a short-hand notation. For example the first equation in Table A1 for a polynomial of 3 is in short-hand notation

$$p = c + \mathrm{poly}\left(\mathrm{bc}(t), 3\right) \tag{A.1}$$

and represents

$$p = c + \gamma_{1,1} t + \gamma_{1,2} t^2 + \gamma_{1,1} t^3. \tag{A.2}$$

Table A1: Description of all tested functional forms for polynomial regression. All functional forms were tested for a maximum polynomial degree from 3 to 13. The shorthand notation was used for the description.

Number	Functional form	Range of γ
1	$p = c + \text{poly}\,(\text{bc}(t), \gamma)$	$\gamma \in \mathbb{Z}\,[3, 13]$
2	$p = c + \text{poly}\,(\text{bc}(t), \gamma) + k$	$\gamma \in \mathbb{Z}\,[3, 13]$
3	$p = c + \text{poly}\,(\text{bc}(t), \gamma) * k$	$\gamma \in \mathbb{Z}\,[3, 13]$
4	$p = c + \text{poly}\,(\text{bc}(t), \gamma) + \log(k)$	$\gamma \in \mathbb{Z}\,[3, 13]$
5	$p = c + \text{poly}\,(\text{bc}(t), \gamma) * \log(k)$	$\gamma \in \mathbb{Z}\,[3, 13]$
6	$p = c + \text{poly}\,(\text{bc}(t), \gamma) + k\_d$	$\gamma \in \mathbb{Z}\,[3, 13]$
7	$p = c + \text{poly}\,(\text{bc}(t), \gamma) * k\_d$	$\gamma \in \mathbb{Z}\,[3, 13]$
8	$\log(p) = c + \text{poly}\,(\text{bc}(t), \gamma)$	$\gamma \in \mathbb{Z}\,[3, 13]$
9	$\log(p) = c + \text{poly}\,(\text{bc}(t), \gamma) + k$	$\gamma \in \mathbb{Z}\,[3, 13]$
10	$\log(p) = c + \text{poly}\,(\text{bc}(t), \gamma) * k$	$\gamma \in \mathbb{Z}\,[3, 13]$
11	$\log(p) = c + \text{poly}\,(\text{bc}(t), \gamma) + \log(k)$	$\gamma \in \mathbb{Z}\,[3, 13]$
12	$\log(p) = c + \text{poly}\,(\text{bc}(t), \gamma) * \log(k)$	$\gamma \in \mathbb{Z}\,[3, 13]$
13	$\log(p) = c + \text{poly}\,(\text{bc}(t), \gamma) + k\_d$	$\gamma \in \mathbb{Z}\,[3, 13]$
14	$\log(p) = c + \text{poly}\,(\text{bc}(t), \gamma) * k\_d$	$\gamma \in \mathbb{Z}\,[3, 13]$
15	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), \gamma)$	$\gamma \in \mathbb{Z}\,[3, 13]$
16	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), \gamma) + k$	$\gamma \in \mathbb{Z}\,[3, 13]$
17	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), \gamma) * k$	$\gamma \in \mathbb{Z}\,[3, 13]$
18	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), \gamma) + \log(k)$	$\gamma \in \mathbb{Z}\,[3, 13]$
19	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), \gamma) * \log(k)$	$\gamma \in \mathbb{Z}\,[3, 13]$
20	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), \gamma) + k\_d$	$\gamma \in \mathbb{Z}\,[3, 13]$
21	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), \gamma) * k\_d$	$\gamma \in \mathbb{Z}\,[3, 13]$

Table A2: The five best models, based on the cRMSE for the lower tail of the distribution, for the first case (no constant, no trend) and all underlying tests included. The RMSE and cRMSE were calculated over the whole distribution and over the lower tail of the distribution. The cRMSE reflects the RMSE after correcting for values ranging between 0 and 1.

		Full Distribution		**Lower Tail** ($p \leq 0.2$)	
	Model	RMSE	cRMSE	RMSE	cRMSE
1	$p = c + \text{poly}\,(\text{bc}(t), 13) * k\_d$	1.79e-04	1.73e-04	1.73e-04	1.71e-04
2	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 13) * k\_d$	1.76e-04	1.74e-04	1.88e-04	1.86e-04
3	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 12) * k\_d$	2.00e-04	1.95e-04	2.10e-04	2.05e-04
4	$p = c + \text{poly}\,(\text{bc}(t), 12) * k\_d$	2.40e-04	2.27e-04	2.28e-04	2.18e-04
5	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 11) * k\_d$	2.16e-04	2.09e-04	2.28e-04	2.19e-04

Table A3: Performance of the models for the first case and all underlying tests included. The RMSE and cRMSE were calculated over the whole distribution and over the lower tail of the distribution. The cRMSE reflects the RMSE after correcting for values ranging between 0 and 1.

		Full Distribution		**Lower Tail** ($p \leq 0.2$)	
	Model	RMSE	cRMSE	RMSE	cRMSE
1	$p = c + \text{poly}\,(\text{bc}(t), 3)$	3.21e-02	2.38e-02	2.51e-02	2.45e-02
2	$p = c + \text{poly}\,(\text{bc}(t), 4)$	2.48e-02	2.40e-02	2.59e-02	2.55e-02
3	$p = c + \text{poly}\,(\text{bc}(t), 5)$	2.23e-02	2.16e-02	2.15e-02	2.15e-02
4	$p = c + \text{poly}\,(\text{bc}(t), 6)$	1.92e-02	1.87e-02	1.92e-02	1.91e-02
5	$p = c + \text{poly}\,(\text{bc}(t), 7)$	1.82e-02	1.78e-02	1.95e-02	1.90e-02
6	$p = c + \text{poly}\,(\text{bc}(t), 8)$	1.68e-02	1.67e-02	1.81e-02	1.81e-02
7	$p = c + \text{poly}\,(\text{bc}(t), 9)$	1.67e-02	1.66e-02	1.82e-02	1.81e-02
8	$p = c + \text{poly}\,(\text{bc}(t), 10)$	1.66e-02	1.66e-02	1.81e-02	1.81e-02
9	$p = c + \text{poly}\,(\text{bc}(t), 11)$	1.65e-02	1.65e-02	1.80e-02	1.80e-02
10	$p = c + \text{poly}\,(\text{bc}(t), 12)$	1.65e-02	1.65e-02	1.80e-02	1.80e-02
11	$p = c + \text{poly}\,(\text{bc}(t), 13)$	1.65e-02	1.65e-02	1.80e-02	1.80e-02
12	$p = c + \text{poly}\,(\text{bc}(t), 3) + k$	3.04e-02	2.11e-02	2.07e-02	1.98e-02
13	$p = c + \text{poly}\,(\text{bc}(t), 4) + k$	2.25e-02	2.14e-02	2.16e-02	2.10e-02
14	$p = c + \text{poly}\,(\text{bc}(t), 5) + k$	1.97e-02	1.86e-02	1.60e-02	1.58e-02

Table A3: Performance of the models for the first case and all underlying tests included. The RMSE and cRMSE were calculated over the whole distribution and over the lower tail of the distribution. The cRMSE reflects the RMSE after correcting for values ranging between 0 and 1. *(continued)*

	Model	RMSE	cRMSE	RMSE	cRMSE
15	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 6) + k$	1.60e-02	1.53e-02	1.27e-02	1.26e-02
16	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 7) + k$	1.49e-02	1.42e-02	1.31e-02	1.24e-02
17	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 8) + k$	1.30e-02	1.29e-02	1.10e-02	1.09e-02
18	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 9) + k$	1.29e-02	1.28e-02	1.10e-02	1.10e-02
19	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 10) + k$	1.28e-02	1.28e-02	1.09e-02	1.09e-02
20	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 11) + k$	1.27e-02	1.26e-02	1.08e-02	1.08e-02
21	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 12) + k$	1.27e-02	1.26e-02	1.08e-02	1.08e-02
22	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 13) + k$	1.27e-02	1.26e-02	1.08e-02	1.08e-02
23	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 3) \ast k$	2.77e-02	1.74e-02	1.82e-02	1.72e-02
24	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 4) \ast k$	1.85e-02	1.74e-02	1.89e-02	1.82e-02
25	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 5) \ast k$	1.42e-02	1.39e-02	1.19e-02	1.18e-02
26	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 6) \ast k$	8.65e-03	7.68e-03	8.52e-03	7.61e-03
27	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 7) \ast k$	6.90e-03	6.06e-03	6.51e-03	6.22e-03
28	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 8) \ast k$	5.41e-03	5.21e-03	5.60e-03	5.55e-03
29	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 9) \ast k$	5.23e-03	5.10e-03	5.55e-03	5.49e-03
30	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 10) \ast k$	4.81e-03	4.79e-03	5.26e-03	5.25e-03
31	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 11) \ast k$	4.79e-03	4.78e-03	5.24e-03	5.23e-03
32	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 12) \ast k$	4.76e-03	4.75e-03	5.22e-03	5.22e-03
33	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 13) \ast k$	4.75e-03	4.75e-03	5.22e-03	5.22e-03
34	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 3) + \log(k)$	3.03e-02	2.07e-02	2.02e-02	1.93e-02
35	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 4) + \log(k)$	2.23e-02	2.10e-02	2.11e-02	2.05e-02
36	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 5) + \log(k)$	1.94e-02	1.82e-02	1.54e-02	1.52e-02
37	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 6) + \log(k)$	1.56e-02	1.48e-02	1.19e-02	1.18e-02
38	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 7) + \log(k)$	1.45e-02	1.38e-02	1.23e-02	1.15e-02
39	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 8) + \log(k)$	1.26e-02	1.24e-02	1.00e-02	1.00e-02
40	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 9) + \log(k)$	1.25e-02	1.23e-02	1.01e-02	1.00e-02

Table A3: Performance of the models for the first case and all underlying tests included. The RMSE and cRMSE were calculated over the whole distribution and over the lower tail of the distribution. The cRMSE reflects the RMSE after correcting for values ranging between 0 and 1. *(continued)*

	Model	RMSE	cRMSE	RMSE	cRMSE
41	$p = c + \text{poly}\left(\text{bc}(t), 10\right) + \log(k)$	1.24e-02	1.23e-02	9.96e-03	9.93e-03
42	$p = c + \text{poly}\left(\text{bc}(t), 11\right) + \log(k)$	1.22e-02	1.21e-02	9.88e-03	9.84e-03
43	$p = c + \text{poly}\left(\text{bc}(t), 12\right) + \log(k)$	1.22e-02	1.21e-02	9.85e-03	9.83e-03
44	$p = c + \text{poly}\left(\text{bc}(t), 13\right) + \log(k)$	1.22e-02	1.21e-02	9.85e-03	9.83e-03
45	$p = c + \text{poly}\left(\text{bc}(t), 3\right) * \log(k)$	2.74e-02	1.69e-02	1.76e-02	1.66e-02
46	$p = c + \text{poly}\left(\text{bc}(t), 4\right) * \log(k)$	1.83e-02	1.70e-02	1.85e-02	1.77e-02
47	$p = c + \text{poly}\left(\text{bc}(t), 5\right) * \log(k)$	1.33e-02	1.31e-02	1.05e-02	1.05e-02
48	$p = c + \text{poly}\left(\text{bc}(t), 6\right) * \log(k)$	6.94e-03	5.74e-03	6.79e-03	5.48e-03
49	$p = c + \text{poly}\left(\text{bc}(t), 7\right) * \log(k)$	5.18e-03	3.91e-03	3.90e-03	3.47e-03
50	$p = c + \text{poly}\left(\text{bc}(t), 8\right) * \log(k)$	2.70e-03	2.29e-03	2.18e-03	2.04e-03
51	$p = c + \text{poly}\left(\text{bc}(t), 9\right) * \log(k)$	2.40e-03	2.08e-03	2.08e-03	1.90e-03
52	$p = c + \text{poly}\left(\text{bc}(t), 10\right) * \log(k)$	1.03e-03	9.71e-04	9.22e-04	8.74e-04
53	$p = c + \text{poly}\left(\text{bc}(t), 11\right) * \log(k)$	1.03e-03	9.68e-04	9.34e-04	8.82e-04
54	$p = c + \text{poly}\left(\text{bc}(t), 12\right) * \log(k)$	8.29e-04	8.02e-04	7.39e-04	7.35e-04
55	$p = c + \text{poly}\left(\text{bc}(t), 13\right) * \log(k)$	7.71e-04	7.63e-04	7.37e-04	7.31e-04
56	$p = c + \text{poly}\left(\text{bc}(t), 3\right) + k\_d$	3.03e-02	2.07e-02	2.02e-02	1.93e-02
57	$p = c + \text{poly}\left(\text{bc}(t), 4\right) + k\_d$	2.23e-02	2.10e-02	2.11e-02	2.05e-02
58	$p = c + \text{poly}\left(\text{bc}(t), 5\right) + k\_d$	1.94e-02	1.82e-02	1.54e-02	1.52e-02
59	$p = c + \text{poly}\left(\text{bc}(t), 6\right) + k\_d$	1.56e-02	1.48e-02	1.18e-02	1.18e-02
60	$p = c + \text{poly}\left(\text{bc}(t), 7\right) + k\_d$	1.45e-02	1.38e-02	1.23e-02	1.15e-02
61	$p = c + \text{poly}\left(\text{bc}(t), 8\right) + k\_d$	1.26e-02	1.24e-02	1.00e-02	1.00e-02
62	$p = c + \text{poly}\left(\text{bc}(t), 9\right) + k\_d$	1.25e-02	1.23e-02	1.01e-02	1.00e-02
63	$p = c + \text{poly}\left(\text{bc}(t), 10\right) + k\_d$	1.24e-02	1.23e-02	9.95e-03	9.92e-03
64	$p = c + \text{poly}\left(\text{bc}(t), 11\right) + k\_d$	1.22e-02	1.21e-02	9.87e-03	9.83e-03
65	$p = c + \text{poly}\left(\text{bc}(t), 12\right) + k\_d$	1.22e-02	1.21e-02	9.85e-03	9.82e-03
66	$p = c + \text{poly}\left(\text{bc}(t), 13\right) + k\_d$	1.22e-02	1.21e-02	9.84e-03	9.82e-03

Table A3: Performance of the models for the first case and all underlying tests included. The RMSE and cRMSE were calculated over the whole distribution and over the lower tail of the distribution. The cRMSE reflects the RMSE after correcting for values ranging between 0 and 1. *(continued)*

	Model	RMSE	cRMSE	RMSE	cRMSE
67	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 3) * k\_d$	2.73e-02	1.69e-02	1.76e-02	1.65e-02
68	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 4) * k\_d$	1.79e-02	1.67e-02	1.82e-02	1.74e-02
69	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 5) * k\_d$	1.31e-02	1.29e-02	1.04e-02	1.04e-02
70	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 6) * k\_d$	6.23e-03	5.18e-03	6.20e-03	4.95e-03
71	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 7) * k\_d$	4.52e-03	3.50e-03	3.43e-03	3.07e-03
72	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 8) * k\_d$	2.28e-03	1.90e-03	1.80e-03	1.66e-03
73	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 9) * k\_d$	2.01e-03	1.71e-03	1.73e-03	1.57e-03
74	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 10) * k\_d$	6.70e-04	6.04e-04	5.69e-04	5.19e-04
75	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 11) * k\_d$	5.22e-04	4.65e-04	4.32e-04	3.90e-04
76	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 12) * k\_d$	2.40e-04	2.27e-04	2.28e-04	2.18e-04
77	$p = c + \mathrm{poly}\,(\mathrm{bc}(t), 13) * k\_d$	1.79e-04	1.73e-04	1.73e-04	1.71e-04
78	$\log(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 3)$	2.77e-02	1.93e-02	3.07e-02	2.11e-02
79	$\log(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 4)$	2.13e-02	2.13e-02	2.34e-02	2.34e-02
80	$\log(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 5)$	1.81e-02	1.70e-02	1.99e-02	1.86e-02
81	$\log(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 6)$	1.76e-02	1.69e-02	1.93e-02	1.84e-02
82	$\log(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 7)$	1.71e-02	1.70e-02	1.87e-02	1.85e-02
83	$\log(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 8)$	4.36e-02	1.72e-02	4.86e-02	1.88e-02
84	$\log(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 9)$	2.18e-02	1.97e-02	2.40e-02	2.16e-02
85	$\log(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 10)$	1.77e+04	2.00e-02	1.98e+04	2.20e-02
86	$\log(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 11)$	1.73e-02	1.70e-02	1.90e-02	1.86e-02
87	$\log(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 12)$	1.66e-02	1.66e-02	1.81e-02	1.81e-02
88	$\log(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 13)$	5.36e+00	1.75e-02	6.00e+00	1.91e-02
89	$\log(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 3) + k$	3.04e-02	2.30e-02	3.38e-02	2.54e-02
90	$\log(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 4) + k$	2.43e-02	2.43e-02	2.69e-02	2.69e-02
91	$\log(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 5) + k$	2.15e-02	2.06e-02	2.38e-02	2.27e-02
92	$\log(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 6) + k$	2.11e-02	2.05e-02	2.33e-02	2.26e-02

Table A3: Performance of the models for the first case and all underlying tests included. The RMSE and cRMSE were calculated over the whole distribution and over the lower tail of the distribution. The cRMSE reflects the RMSE after correcting for values ranging between 0 and 1. *(continued)*

	Model	RMSE	cRMSE	RMSE	cRMSE
93	$\log(p) = c + \text{poly}\left(\text{bc}(t), 7\right) + k$	2.06e-02	2.05e-02	2.28e-02	2.26e-02
94	$\log(p) = c + \text{poly}\left(\text{bc}(t), 8\right) + k$	4.54e-02	2.08e-02	5.06e-02	2.29e-02
95	$\log(p) = c + \text{poly}\left(\text{bc}(t), 9\right) + k$	2.48e-02	2.29e-02	2.74e-02	2.53e-02
96	$\log(p) = c + \text{poly}\left(\text{bc}(t), 10\right) + k$	1.78e+04	2.31e-02	1.99e+04	2.56e-02
97	$\log(p) = c + \text{poly}\left(\text{bc}(t), 11\right) + k$	2.09e-02	2.06e-02	2.31e-02	2.28e-02
98	$\log(p) = c + \text{poly}\left(\text{bc}(t), 12\right) + k$	2.03e-02	2.03e-02	2.24e-02	2.24e-02
99	$\log(p) = c + \text{poly}\left(\text{bc}(t), 13\right) + k$	5.37e+00	2.10e-02	6.00e+00	2.32e-02
100	$\log(p) = c + \text{poly}\left(\text{bc}(t), 3\right) * k$	2.85e-02	1.37e-02	3.18e-02	1.53e-02
101	$\log(p) = c + \text{poly}\left(\text{bc}(t), 4\right) * k$	1.13e-02	1.13e-02	1.26e-02	1.26e-02
102	$\log(p) = c + \text{poly}\left(\text{bc}(t), 5\right) * k$	7.47e-03	5.87e-03	8.30e-03	6.50e-03
103	$\log(p) = c + \text{poly}\left(\text{bc}(t), 6\right) * k$	8.95e-03	8.11e-03	9.97e-03	9.02e-03
104	$\log(p) = c + \text{poly}\left(\text{bc}(t), 7\right) * k$	1.97e+05	1.67e-02	2.20e+05	1.86e-02
105	$\log(p) = c + \text{poly}\left(\text{bc}(t), 8\right) * k$	1.59e-02	1.20e-02	1.78e-02	1.34e-02
106	$\log(p) = c + \text{poly}\left(\text{bc}(t), 9\right) * k$	1.04e-01	6.10e-03	1.16e-01	6.78e-03
107	$\log(p) = c + \text{poly}\left(\text{bc}(t), 10\right) * k$	5.25e-03	5.12e-03	5.81e-03	5.67e-03
108	$\log(p) = c + \text{poly}\left(\text{bc}(t), 11\right) * k$	6.85e-03	5.94e-03	7.61e-03	6.59e-03
109	$\log(p) = c + \text{poly}\left(\text{bc}(t), 12\right) * k$	1.64e+00	7.03e-03	1.83e+00	7.80e-03
110	$\log(p) = c + \text{poly}\left(\text{bc}(t), 13\right) * k$	2.36e+02	5.27e-03	6.22e-03	5.82e-03
111	$\log(p) = c + \text{poly}\left(\text{bc}(t), 3\right) + \log(k)$	3.07e-02	2.33e-02	3.41e-02	2.58e-02
112	$\log(p) = c + \text{poly}\left(\text{bc}(t), 4\right) + \log(k)$	2.46e-02	2.45e-02	2.72e-02	2.72e-02
113	$\log(p) = c + \text{poly}\left(\text{bc}(t), 5\right) + \log(k)$	2.18e-02	2.09e-02	2.41e-02	2.31e-02
114	$\log(p) = c + \text{poly}\left(\text{bc}(t), 6\right) + \log(k)$	2.14e-02	2.08e-02	2.37e-02	2.30e-02
115	$\log(p) = c + \text{poly}\left(\text{bc}(t), 7\right) + \log(k)$	2.10e-02	2.08e-02	2.31e-02	2.30e-02
116	$\log(p) = c + \text{poly}\left(\text{bc}(t), 8\right) + \log(k)$	4.58e-02	2.11e-02	5.11e-02	2.33e-02
117	$\log(p) = c + \text{poly}\left(\text{bc}(t), 9\right) + \log(k)$	2.50e-02	2.32e-02	2.77e-02	2.57e-02
118	$\log(p) = c + \text{poly}\left(\text{bc}(t), 10\right) + \log(k)$	1.78e+04	2.34e-02	1.98e+04	2.59e-02

Table A3: Performance of the models for the first case and all underlying tests included. The RMSE and cRMSE were calculated over the whole distribution and over the lower tail of the distribution. The cRMSE reflects the RMSE after correcting for values ranging between 0 and 1. *(continued)*

	Model	RMSE	cRMSE	RMSE	cRMSE
119	$\log(p) = c + \text{poly}\left(\text{bc}(t), 11\right) + \log(k)$	2.12e-02	2.09e-02	2.34e-02	2.31e-02
120	$\log(p) = c + \text{poly}\left(\text{bc}(t), 12\right) + \log(k)$	2.06e-02	2.06e-02	2.27e-02	2.27e-02
121	$\log(p) = c + \text{poly}\left(\text{bc}(t), 13\right) + \log(k)$	5.33e+00	2.13e-02	5.96e+00	2.35e-02
122	$\log(p) = c + \text{poly}\left(\text{bc}(t), 3\right) * \log(k)$	2.88e-02	1.32e-02	3.21e-02	1.47e-02
123	$\log(p) = c + \text{poly}\left(\text{bc}(t), 4\right) * \log(k)$	9.50e-03	9.49e-03	1.06e-02	1.06e-02
124	$\log(p) = c + \text{poly}\left(\text{bc}(t), 5\right) * \log(k)$	7.11e-03	4.01e-03	7.91e-03	4.42e-03
125	$\log(p) = c + \text{poly}\left(\text{bc}(t), 6\right) * \log(k)$	7.80e-03	6.89e-03	8.68e-03	7.66e-03
126	$\log(p) = c + \text{poly}\left(\text{bc}(t), 7\right) * \log(k)$	2.44e+03	1.62e-02	2.73e+03	1.81e-02
127	$\log(p) = c + \text{poly}\left(\text{bc}(t), 8\right) * \log(k)$	1.32e-02	9.75e-03	1.47e-02	1.08e-02
128	$\log(p) = c + \text{poly}\left(\text{bc}(t), 9\right) * \log(k)$	1.15e-02	2.94e-03	1.29e-02	3.22e-03
129	$\log(p) = c + \text{poly}\left(\text{bc}(t), 10\right) * \log(k)$	3.82e-03	1.96e-03	4.22e-03	2.11e-03
130	$\log(p) = c + \text{poly}\left(\text{bc}(t), 11\right) * \log(k)$	8.76e-03	5.56e-03	9.75e-03	6.16e-03
131	$\log(p) = c + \text{poly}\left(\text{bc}(t), 12\right) * \log(k)$	2.22e+01	6.42e-03	2.48e+01	7.13e-03
132	$\log(p) = c + \text{poly}\left(\text{bc}(t), 13\right) * \log(k)$	3.48e+00	2.21e-03	3.48e-03	2.34e-03
133	$\log(p) = c + \text{poly}\left(\text{bc}(t), 3\right) + k\_d$	3.07e-02	2.33e-02	3.41e-02	2.58e-02
134	$\log(p) = c + \text{poly}\left(\text{bc}(t), 4\right) + k\_d$	2.46e-02	2.45e-02	2.72e-02	2.72e-02
135	$\log(p) = c + \text{poly}\left(\text{bc}(t), 5\right) + k\_d$	2.18e-02	2.09e-02	2.41e-02	2.31e-02
136	$\log(p) = c + \text{poly}\left(\text{bc}(t), 6\right) + k\_d$	2.14e-02	2.08e-02	2.37e-02	2.30e-02
137	$\log(p) = c + \text{poly}\left(\text{bc}(t), 7\right) + k\_d$	2.09e-02	2.08e-02	2.31e-02	2.30e-02
138	$\log(p) = c + \text{poly}\left(\text{bc}(t), 8\right) + k\_d$	4.58e-02	2.11e-02	5.10e-02	2.33e-02
139	$\log(p) = c + \text{poly}\left(\text{bc}(t), 9\right) + k\_d$	2.50e-02	2.32e-02	2.77e-02	2.57e-02
140	$\log(p) = c + \text{poly}\left(\text{bc}(t), 10\right) + k\_d$	1.78e+04	2.34e-02	1.98e+04	2.59e-02
141	$\log(p) = c + \text{poly}\left(\text{bc}(t), 11\right) + k\_d$	2.12e-02	2.09e-02	2.34e-02	2.31e-02
142	$\log(p) = c + \text{poly}\left(\text{bc}(t), 12\right) + k\_d$	2.06e-02	2.06e-02	2.27e-02	2.27e-02
143	$\log(p) = c + \text{poly}\left(\text{bc}(t), 13\right) + k\_d$	5.34e+00	2.13e-02	5.97e+00	2.35e-02
144	$\log(p) = c + \text{poly}\left(\text{bc}(t), 3\right) * k\_d$	2.82e-02	1.32e-02	3.15e-02	1.47e-02

Table A3: Performance of the models for the first case and all underlying tests included. The RMSE and cRMSE were calculated over the whole distribution and over the lower tail of the distribution. The cRMSE reflects the RMSE after correcting for values ranging between 0 and 1. *(continued)*

	Model	RMSE	cRMSE	RMSE	cRMSE
145	$\log(p) = c + \text{poly}\,(\text{bc}(t), 4) * k\_d$	9.75e-03	9.72e-03	1.09e-02	1.08e-02
146	$\log(p) = c + \text{poly}\,(\text{bc}(t), 5) * k\_d$	5.96e-03	3.23e-03	6.65e-03	3.59e-03
147	$\log(p) = c + \text{poly}\,(\text{bc}(t), 6) * k\_d$	8.80e-03	7.54e-03	9.82e-03	8.42e-03
148	$\log(p) = c + \text{poly}\,(\text{bc}(t), 7) * k\_d$	4.24e+05	1.91e-02	4.74e+05	2.12e-02
149	$\log(p) = c + \text{poly}\,(\text{bc}(t), 8) * k\_d$	2.82e-02	1.74e-02	2.78e-02	1.94e-02
150	$\log(p) = c + \text{poly}\,(\text{bc}(t), 9) * k\_d$	3.17e+01	1.23e-02	3.54e+01	1.36e-02
151	$\log(p) = c + \text{poly}\,(\text{bc}(t), 10) * k\_d$	1.68e-02	7.51e-03	1.87e-02	8.35e-03
152	$\log(p) = c + \text{poly}\,(\text{bc}(t), 11) * k\_d$	8.20e-03	2.40e-03	9.17e-03	2.68e-03
153	$\log(p) = c + \text{poly}\,(\text{bc}(t), 12) * k\_d$	7.66e-04	6.42e-04	8.52e-04	7.13e-04
154	$\log(p) = c + \text{poly}\,(\text{bc}(t), 13) * k\_d$	3.38e-04	3.05e-04	3.72e-04	3.34e-04
155	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 3)$	4.22e-02	2.28e-02	2.47e-02	2.44e-02
156	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 4)$	2.68e-02	2.53e-02	2.77e-02	2.75e-02
157	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 5)$	1.87e-02	1.82e-02	1.95e-02	1.95e-02
158	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 6)$	1.75e-02	1.72e-02	1.90e-02	1.86e-02
159	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 7)$	1.76e-02	1.72e-02	1.91e-02	1.87e-02
160	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 8)$	1.66e-02	1.66e-02	1.81e-02	1.81e-02
161	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 9)$	1.66e-02	1.66e-02	1.81e-02	1.81e-02
162	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 10)$	1.65e-02	1.65e-02	1.80e-02	1.80e-02
163	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 11)$	1.65e-02	1.65e-02	1.80e-02	1.80e-02
164	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 12)$	1.65e-02	1.65e-02	1.80e-02	1.80e-02
165	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 13)$	1.65e-02	1.65e-02	1.80e-02	1.80e-02
166	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 3) + k$	4.09e-02	1.95e-02	2.04e-02	1.99e-02
167	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 4) + k$	2.42e-02	2.24e-02	2.40e-02	2.35e-02
168	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 5) + k$	1.46e-02	1.38e-02	1.34e-02	1.34e-02
169	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 6) + k$	1.30e-02	1.25e-02	1.27e-02	1.21e-02
170	$\text{bc}(p) = c + \text{poly}\,(\text{bc}(t), 7) + k$	1.30e-02	1.25e-02	1.28e-02	1.21e-02

Table A3: Performance of the models for the first case and all underlying tests included. The RMSE and cRMSE were calculated over the whole distribution and over the lower tail of the distribution. The cRMSE reflects the RMSE after correcting for values ranging between 0 and 1. *(continued)*

	Model	RMSE	cRMSE	RMSE	cRMSE
171	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 8) + k$	1.16e-02	1.16e-02	1.12e-02	1.12e-02
172	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 9) + k$	1.16e-02	1.16e-02	1.12e-02	1.12e-02
173	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 10) + k$	1.15e-02	1.15e-02	1.11e-02	1.11e-02
174	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 11) + k$	1.15e-02	1.15e-02	1.11e-02	1.11e-02
175	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 12) + k$	1.15e-02	1.15e-02	1.11e-02	1.11e-02
176	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 13) + k$	1.15e-02	1.15e-02	1.11e-02	1.11e-02
177	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 3) * k$	3.74e-02	1.56e-02	1.71e-02	1.65e-02
178	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 4) * k$	2.05e-02	1.89e-02	2.09e-02	2.04e-02
179	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 5) * k$	8.16e-03	7.98e-03	8.24e-03	8.18e-03
180	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 6) * k$	7.62e-03	6.52e-03	8.27e-03	7.01e-03
181	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 7) * k$	5.59e-03	5.27e-03	5.88e-03	5.71e-03
182	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 8) * k$	5.09e-03	5.02e-03	5.60e-03	5.52e-03
183	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 9) * k$	4.86e-03	4.85e-03	5.36e-03	5.34e-03
184	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 10) * k$	4.82e-03	4.81e-03	5.31e-03	5.30e-03
185	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 11) * k$	4.80e-03	4.80e-03	5.28e-03	5.28e-03
186	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 12) * k$	4.79e-03	4.79e-03	5.28e-03	5.28e-03
187	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 13) * k$	4.79e-03	4.79e-03	5.28e-03	5.28e-03
188	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 3) + \log(k)$	4.08e-02	1.91e-02	1.99e-02	1.94e-02
189	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 4) + \log(k)$	2.39e-02	2.21e-02	2.36e-02	2.31e-02
190	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 5) + \log(k)$	1.41e-02	1.33e-02	1.27e-02	1.26e-02
191	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 6) + \log(k)$	1.24e-02	1.19e-02	1.19e-02	1.13e-02
192	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 7) + \log(k)$	1.25e-02	1.19e-02	1.20e-02	1.13e-02
193	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 8) + \log(k)$	1.10e-02	1.10e-02	1.04e-02	1.03e-02
194	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 9) + \log(k)$	1.10e-02	1.09e-02	1.03e-02	1.03e-02
195	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 10) + \log(k)$	1.09e-02	1.09e-02	1.02e-02	1.02e-02
196	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 11) + \log(k)$	1.09e-02	1.09e-02	1.02e-02	1.02e-02

Table A3: Performance of the models for the first case and all underlying tests included. The RMSE and cRMSE were calculated over the whole distribution and over the lower tail of the distribution. The cRMSE reflects the RMSE after correcting for values ranging between 0 and 1. *(continued)*

	Model	RMSE	cRMSE	RMSE	cRMSE
197	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 12\right) + \log(k)$	1.09e-02	1.09e-02	1.02e-02	1.02e-02
198	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 13\right) + \log(k)$	1.09e-02	1.09e-02	1.02e-02	1.02e-02
199	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 3\right) * \log(k)$	3.67e-02	1.51e-02	1.66e-02	1.59e-02
200	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 4\right) * \log(k)$	2.03e-02	1.84e-02	2.04e-02	1.99e-02
201	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 5\right) * \log(k)$	6.30e-03	6.22e-03	6.13e-03	6.04e-03
202	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 6\right) * \log(k)$	6.15e-03	4.61e-03	6.63e-03	4.82e-03
203	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 7\right) * \log(k)$	2.93e-03	2.14e-03	2.31e-03	2.06e-03
204	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 8\right) * \log(k)$	1.95e-03	1.76e-03	2.07e-03	1.85e-03
205	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 9\right) * \log(k)$	1.11e-03	1.05e-03	1.11e-03	1.04e-03
206	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 10\right) * \log(k)$	8.95e-04	8.39e-04	8.76e-04	8.04e-04
207	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 11\right) * \log(k)$	7.74e-04	7.62e-04	7.18e-04	7.01e-04
208	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 12\right) * \log(k)$	7.22e-04	7.19e-04	6.67e-04	6.64e-04
209	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 13\right) * \log(k)$	7.12e-04	7.11e-04	6.54e-04	6.53e-04
210	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 3\right) + k\_d$	4.08e-02	1.91e-02	1.99e-02	1.94e-02
211	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 4\right) + k\_d$	2.39e-02	2.21e-02	2.36e-02	2.31e-02
212	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 5\right) + k\_d$	1.41e-02	1.33e-02	1.27e-02	1.26e-02
213	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 6\right) + k\_d$	1.24e-02	1.19e-02	1.19e-02	1.13e-02
214	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 7\right) + k\_d$	1.25e-02	1.19e-02	1.20e-02	1.13e-02
215	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 8\right) + k\_d$	1.10e-02	1.10e-02	1.04e-02	1.03e-02
216	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 9\right) + k\_d$	1.10e-02	1.09e-02	1.03e-02	1.03e-02
217	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 10\right) + k\_d$	1.09e-02	1.09e-02	1.02e-02	1.02e-02
218	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 11\right) + k\_d$	1.09e-02	1.09e-02	1.02e-02	1.02e-02
219	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 12\right) + k\_d$	1.09e-02	1.09e-02	1.02e-02	1.02e-02
220	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 13\right) + k\_d$	1.09e-02	1.09e-02	1.02e-02	1.02e-02
221	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 3\right) * k\_d$	3.67e-02	1.50e-02	1.64e-02	1.58e-02
222	$\text{bc}(p) = c + \text{poly}\left(\text{bc}(t), 4\right) * k\_d$	2.00e-02	1.82e-02	2.01e-02	1.96e-02

Table A3: Performance of the models for the first case and all underlying tests included. The RMSE and cRMSE were calculated over the whole distribution and over the lower tail of the distribution. The cRMSE reflects the RMSE after correcting for values ranging between 0 and 1. *(continued)*

	Model	RMSE	cRMSE	RMSE	cRMSE
223	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 5) * k\_d$	6.14e-03	6.09e-03	6.00e-03	5.93e-03
224	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 6) * k\_d$	5.75e-03	4.29e-03	6.20e-03	4.49e-03
225	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 7) * k\_d$	2.24e-03	1.83e-03	1.96e-03	1.74e-03
226	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 8) * k\_d$	1.54e-03	1.36e-03	1.67e-03	1.45e-03
227	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 9) * k\_d$	7.92e-04	7.23e-04	8.39e-04	7.60e-04
228	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 10) * k\_d$	4.80e-04	4.28e-04	5.13e-04	4.52e-04
229	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 11) * k\_d$	2.16e-04	2.09e-04	2.28e-04	2.19e-04
230	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 12) * k\_d$	2.00e-04	1.95e-04	2.10e-04	2.05e-04
231	$\mathrm{bc}(p) = c + \mathrm{poly}\,(\mathrm{bc}(t), 13) * k\_d$	1.76e-04	1.74e-04	1.88e-04	1.86e-04

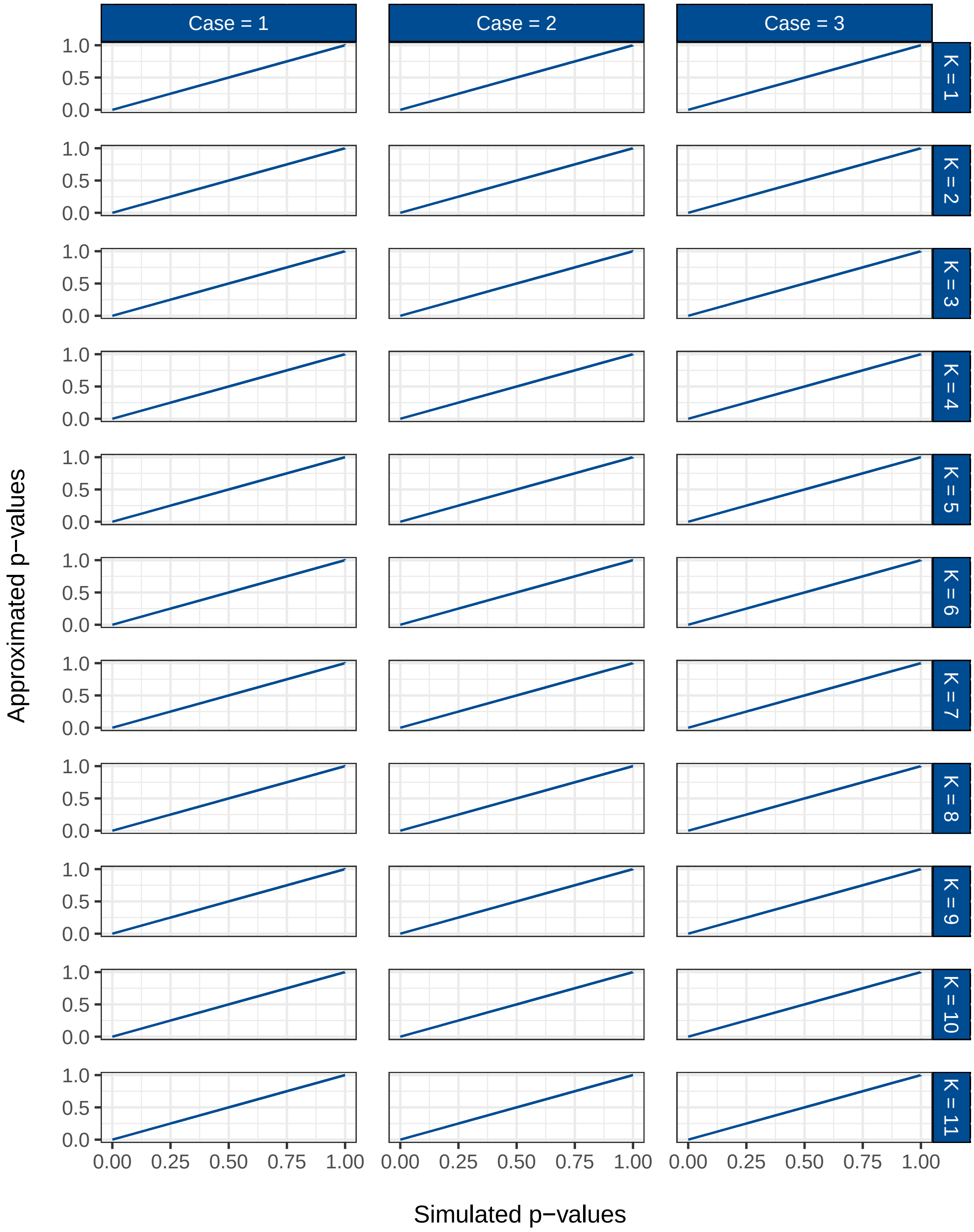

Figure A1: Simulated against approximated p-values over the whole distribution for all cases and all underlying tests.

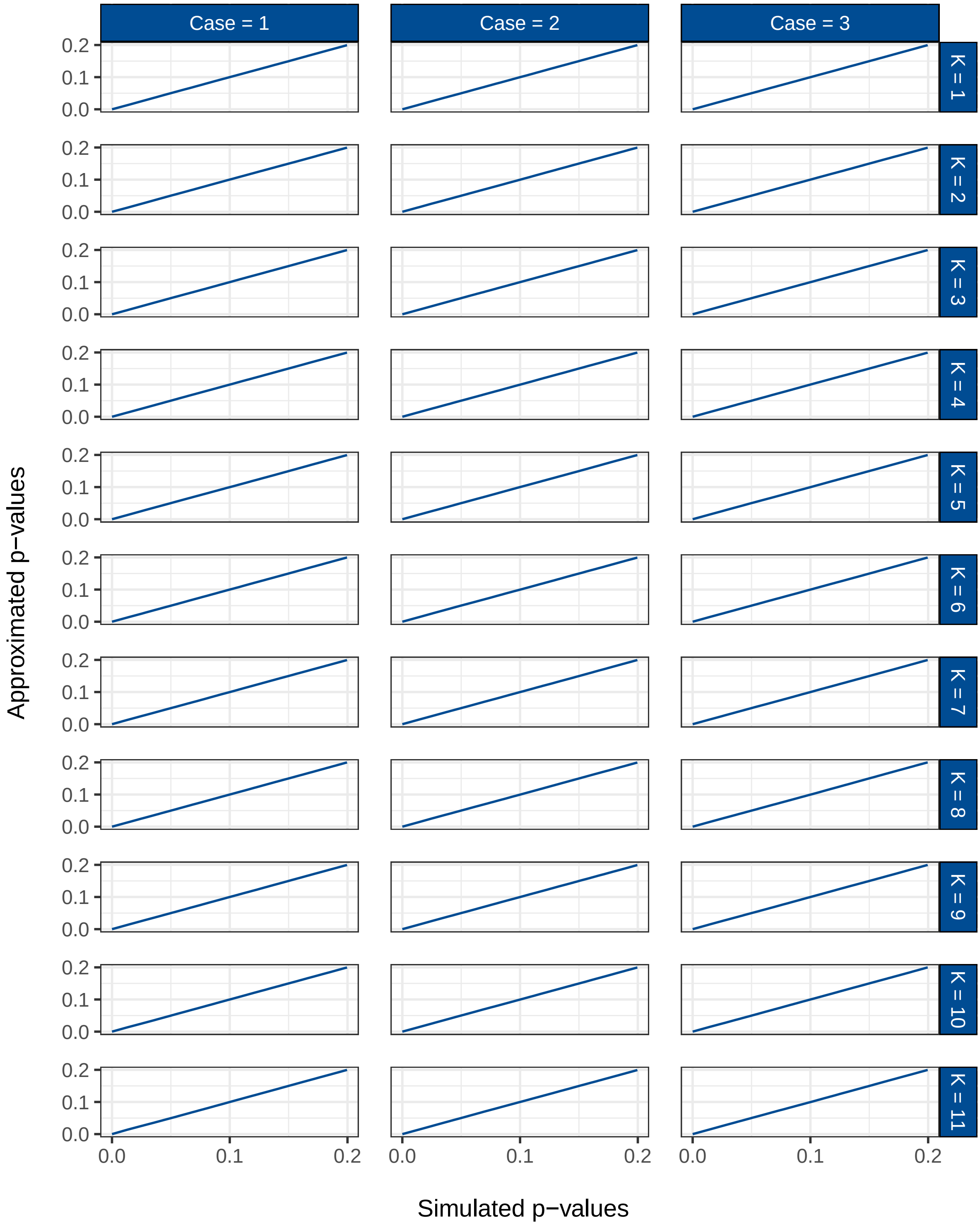

Figure A2: Simulated vs. approximated p-values for the lower tail of the distribution for all cases and all underlying test.

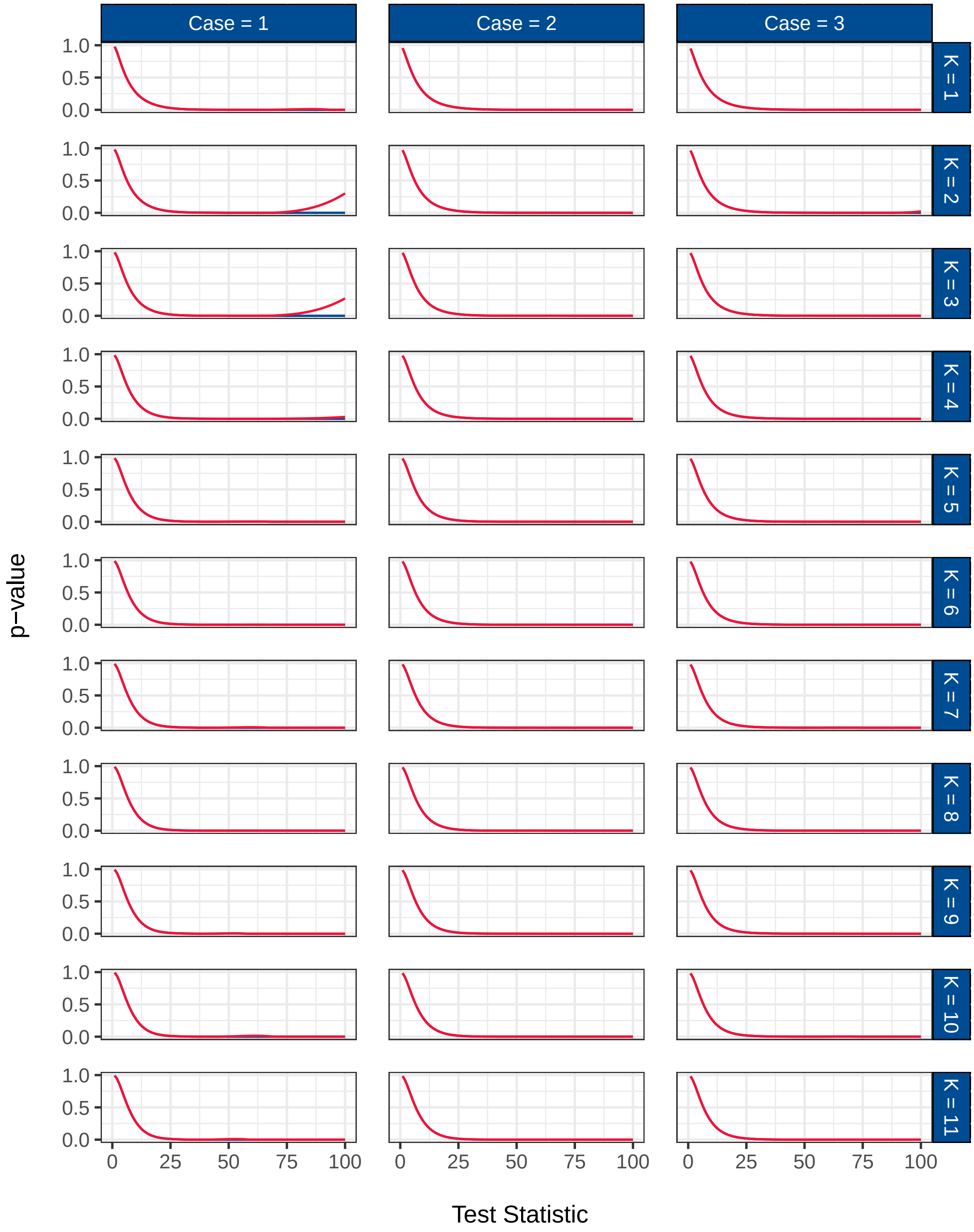

Figure A3: Corrected (blue) and uncorrected (red) p-value predictions for all cases and all underlying tests.

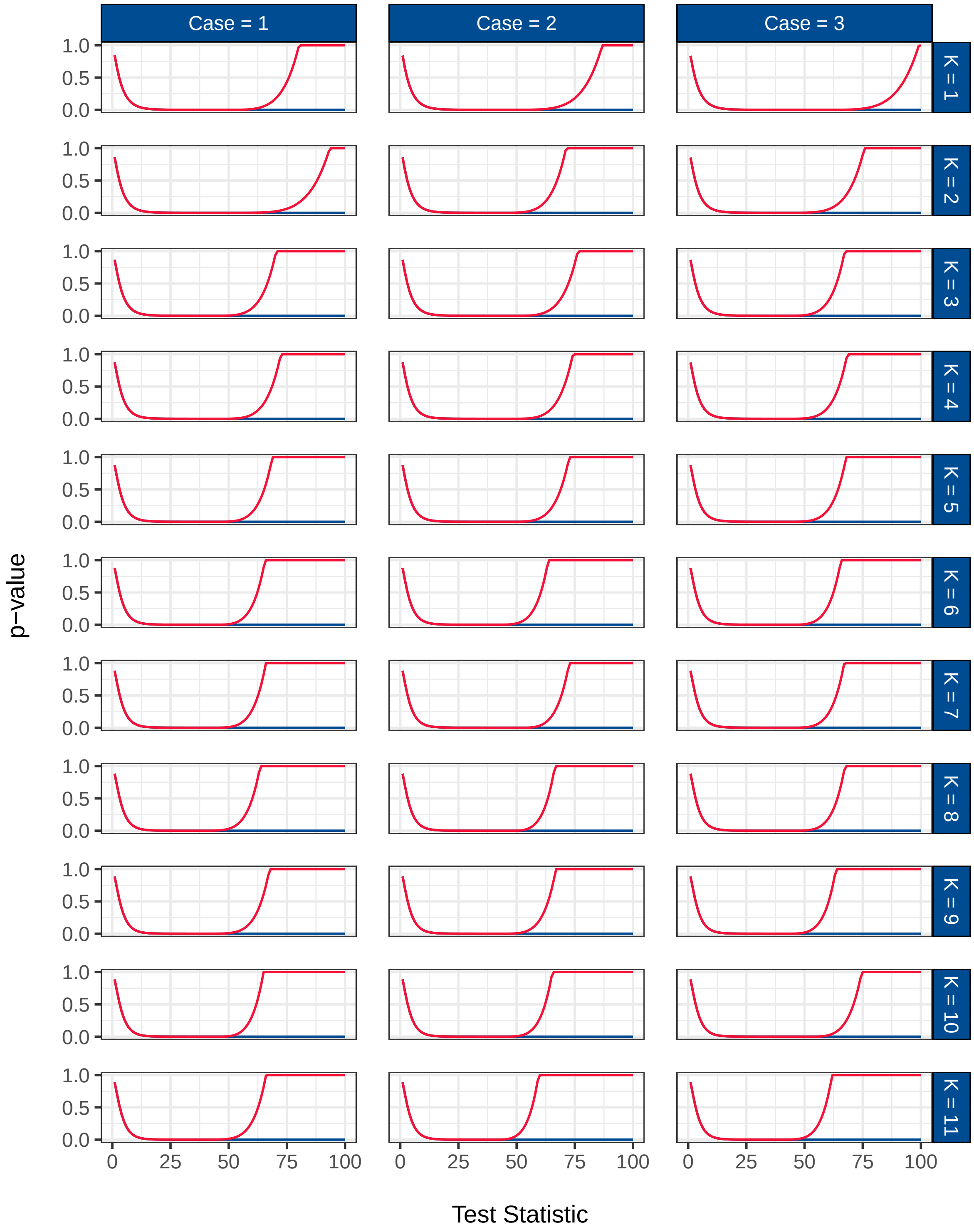

Figure A4: Corrected (blue) and uncorrected (red) p-value predictions for all cases using Engle-Granger and Johansen as underlying tests.

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