

Code documentation

Summary

This document describes the R code that can be used to replicate the empirical results reported in the paper *Multiscale inference for nonparametric time trends*. The overall structure of the code is as follows. There are four main files each of which produces a specific part of the simulations and applications:

`main_shape_simulation.R` produces the simulation results for the test of the hypothesis $H_0 : m' = 0$ from Section 4.

`main_equality_simulation.R` produces the simulation results for the test of the hypothesis $H_0 : m_1 = \dots = m_n$ from Section 5.

`main_shape_application.R` produces the application results from Section 8.1, where the test of the hypothesis $H_0 : m' = 0$ is used.

`main_equality_application.R` produces the application results from Section 8.2, where the test of the hypothesis $H_0 : m_1 = \dots = m_n$ is used.

These main files read in a number of functions which are collected in the two folders **Shape** and **Equality**. The folder **Shape** contains the code for the test methods from Section 4 and **Equality** the code for the methods from Section 5. The simulation and application results are stored either as figures or as `.tex` files (for tables) in the folder **Plots**.

All programs are written in R with some functions in C. They are all quite self-explanatory and commented. The code is self-sufficient, the only thing that is OS-specific and may need to be changed is the extension of dynamic libraries. Dynamic libraries are defined in the beginning of each main file. At the moment everything is configured for running on Windows, that is, the `.dll` extension is used. In order to make the programs Linux or MacOS compatible, this extension should be changed to `.so`. For example,

```
dyn.load("Shape/C_code/estimating_sigma.dll")
```

should be replaced by

```
dyn.load("Shape/C_code/estimating_sigma.so").
```

All the dynamic libraries are already provided, additional compilation is not necessary.

Code for the test methods from Section 4

`main_shape_simulation.R` produces the simulation results for the test method from Section 4, in particular Tables 1 and 2. If you use Linux or MacOS, please change the extension of dynamic libraries in the very beginning of the main file to `.so`. The size and power results in Tables 1 and 2 are calculated from 1000 simulated samples. This number can be set as `N_rep` in `main_shape_simulation.R`. The running time for 1000 simulations is quite small, amounting to ??.

`main_shape_application.R` produces the application results from Section 8.1, in particular Figure 1. As above, please change the extension of dynamic libraries if you use Linux or MacOS.

`Shape` directory containing all the necessary functions, data and simulated distributions to run `main_shape_simulation.R` and `main_shape_application.R`.

`Shape\C_code` folder containing the dynamic libraries and R-wrappers of the C functions.

`Shape\functions.R` folder containing auxiliary R functions.

`Shape\data` folder with the file `cetml1659on.dat` which contains the monthly and yearly mean central England temperature from 1659 up to 2017. It was downloaded from <https://www.metoffice.gov.uk/hadobs/hadcet/data/download.html>. A full description of the data can be accessed on the website of the UK Met Office <https://www.metoffice.gov.uk>.

`Shape\distribution` folder containing simulated distributions of the Gaussian statistic Φ'_T for various time series lengths T . In order to recalculate these distributions, the stored files need to be deleted or removed; the computation is then done automatically. The distribution of the Gaussian statistic Φ'_T used to calculate the critical values of the test in Section 8.1 is stored in `Shape\distribution\distr_T_359_testing_constant_type_11.RData`. As before, the distribution of Φ'_T can be re-calculated by removing or deleting this file before running `N_rep` in `main_equality_simulation.R`.

Code for the test methods from Section 5

`main_equality_simulation.R` produces the simulation results for the methods from Section 5, in particular Tables 3, 4 and 5. If you use Linux or MacOS, please change the extension of dynamic libraries in the very beginning of the main file. The size and power results in Tables 3 and 4 are calculated from 1000 simulated samples. This number can be set as `N_rep` in `main_equality_simulation.R`, though rising this parameter is not advisable due to computational tractability.

`main_equality_application.R` produces the application results from Section 8.2. As above, please change the extension of dynamic libraries if you use Linux or MacOS.

`Equality` directory containing all the necessary functions, data and simulated distributions to run `main_equality_simulation.R` and `main_equality_application.R`.

`Equality\C_code` folder containing the dynamic libraries and R-wrappers of the C functions.

`Equality\functions.R` folder containing auxiliary R functions.

`Equality\data` folder containing historic station data downloaded from the website <https://www.metoffice.gov.uk/public/weather/climate-historic/#?tab=climateHistoric> and the description file `.\Equality\Description_file.xlsx`. The data consist of:

- mean daily maximum temperature (t_{max})
- mean daily minimum temperature (t_{min})
- days of air frost (af)
- total rainfall ($rain$)
- total sunshine duration (sun).

The monthly mean temperature used in the paper is calculated from the average of the mean daily maximum and mean daily minimum temperature, i.e. $(t_{max} + t_{min})/2$. A full description of the data can be accessed on the official website of The UK Met Office <https://www.metoffice.gov.uk>.

`Equality\distribution` folder containing the simulated distributions of the Gaussian statistic $\Phi_{n,T}$ for $n = 15$ time series (the number that was used for the simulation study) and different time series lengths T . If you want to recalculate these distributions, the stored files need to be deleted or removed. The recomputation is then done automatically. However, note that with increasing time series length T , the running time increases significantly. For example, for $T = 1000$ and $n = 15$, the calculation of $\Phi_{n,T}$ may take a few hours. The distribution of the Gaussian statistic $\Phi'_{n,T}$ that was used for calculating quantiles for the application in Section 8.2 is stored in the file `.\Equality\distribution\distr_for_application_T_386_and_N_ts_25_and_method_11.RData`.

Simulations

Testing for the presence of a time trend

The file `.\main_shape_simulation.R` produces the size and power results in Tables 1 and 2 of the paper. As was already mentioned, the only possible adjustment that needs to be done is the change of extension of dynamic libraries in the beginning of `.\main_shape_simulation.R` (if Linux or MacOS is used). Size and power of the test are calculated from 1000 simulated samples. This number can be set as `N_rep` in `.\main_shape_simulation.R`. We have run our simulations for 10000 samples as well, and the running time is quite small.

The folder `.\Shape` contains all the necessary functions, data and simulated distributions to run `.\main_shape_simulation.R`.

The simulated distributions of the Gaussian statistic Φ'_T for various time series lengths T can be found in the directory `.\Shape\distribution`. In order to be able to recalculate these distributions, the stored files need to be deleted or removed; the computation is then done automatically.

The dynamic libraries and R-wrappers of the C functions are provided in the folder `.\Shape\C_code`. All the auxiliary R functions can be found in `.\Shape\functions.R`.

Testing for equality of time trends

In order to obtain Tables 3, 4 and 5 of the paper, one needs to run `.\main_equality_simulation.R`. Do not forget to change the extensions of dynamic libraries in the very beginning of `.\main_equality_simulation.R` if you use Linux or MacOS. Size and power of the test are calculated from 1000 simulated samples. This number can be set as `N_rep` in `.\main_equality_simulation.R`, though rising this parameter is not advisable due to computational tractability.

The folder `.\Equality` contains all the necessary functions, data and simulated distributions required to run `.\main_equality_simulation.R`.

The directory `.\Equality\distribution` comprises the simulated distributions of the Gaussian statistic $\Phi_{n,T}$ for $n = 15$ time series (the number that was used for the simulation study) and different time series lengths T . If you want to recalculate the distribution, the stored files need to be deleted or removed. The recomputation is then done automatically. However, note that with increasing values of time series length T the running time increases significantly. For example, for $T = 1000$ and $n = 15$, the calculation of $\Phi_{n,T}$ may take a few hours.

The dynamic libraries and R-wrappers of the C functions can be found in the folder `.\Equality\C_code`. All the auxiliary R functions are in `.\Equality\functions.R`.

Application

Testing for the presence of a time trend

`.\main_shape_application.R` produces the application results from Section 8.1.

The folder `.\Shape` contains all the necessary functions, data and simulated distributions to run `.\main_shape_application.R`.

The file `.\Shape\data\cetml1659on.dat` contains the monthly and yearly mean central England temperature from the year 1659 up to the year 2017. It was downloaded from <https://www.metoffice.gov.uk/hadobs/hadcet/data/download.html>. A full description of the data can be accessed on the official website of the UK Met Office <https://www.metoffice.gov.uk>.

The distribution of the Gaussian statistic Φ'_T used to calculate the critical values of the test is stored in the file `.\Shape\distribution\distr_T_359_testing_constant_type_11.RData`. As before, the distribution of Φ'_T can be re-calculated by removing or deleting this file before running `N_rep` in `.\main_equality_simulation.R`.

Testing for equality of time trends

`.\main_equality_application.R` produces the application results from Section 8.2.

The folder `.\Equality` contains all the necessary functions, data and simulated distributions to run `.\main_equality_application.R`.

The directory `.\Equality\data` contains the historic station data in `.txt` format downloaded from <https://www.metoffice.gov.uk/public/weather/climate-historic/#?tab=climateHistoric> and the description file `.\Equality\Description_file.xlsx`. The data consists of:

- Mean daily maximum temperature (t_{max})
- Mean daily minimum temperature (t_{min})
- Days of air frost (af)
- Total rainfall ($rain$)
- Total sunshine duration (sun)

The monthly mean temperature that was used in the paper is calculated from the average of the mean daily maximum and mean daily minimum temperature, i.e. $(t_{max} + t_{min})/2$. A full description of the data can be accessed on the official website of The UK Met Office <https://www.metoffice.gov.uk>.

The distribution of the Gaussian statistic $\Phi'_{n,T}$ that was used for calculating quantiles for this application is stored in the file `.\Equality\distribution\distr_for_application_T_386_and_N_ts_25_and_method_11.RData`.