

# Code documentation

## Summary

This document describes the R code that can be used to replicate the empirical results reported in the paper *Multiscale Inference and Long-Run Variance Estimation in Nonparametric Regression with Time Series Errors*. 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 our multiscale method, that are reported in Section 5.1.

`main_shape_comparison.R` produces the simulation results on comparison between our multiscale test and SiZer, that we report in Section 5.2.

`main_variance_simulation.R` produces the simulation results for our long-run variance estimator and its comparison with the estimator from Hall and Van Keilegom (2003) and the oracle one. These results are presented in Section 5.3.

`main_shape_application.R` produces the application results from Section 6, where our multiscale test of the hypothesis  $H_0 : m' = 0$  is applied.

These main files read in a number of functions which are collected in the folder **Shape**. 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, the file

```
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.

## Description of the files provided

`main_shape_simulation.R` produces the simulation results for the test method from Section 3, in particular Tables 1 and 2 in Section 5.1. 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` in `main_shape_simulation.R`. The running time for 1000 simulations is quite small, amounting to a few minutes.

`main_shape_comparison.R` produces the simulation results for comparison between our multiscale method and SiZer from Section 5.2, in particular Tables 3 and 4 and Figure 2. As above, please change the extension of dynamic libraries if you use Linux or MacOS.

`main_variance_simulation.R` produces the simulation results for the long-run variance estimator from Section 4, in particular Figures 3, 4, 5 and 6 from Section 5.3. As above, please change the extension of dynamic libraries if you use Linux or MacOS.

`main_shape_application.R` produces the application results from Section 6, in particular Figure 7. As above, please change the extension of dynamic libraries if you use Linux or MacOS.

`Shape` is a directory that contains all the necessary functions, data and simulated distributions to run the four main files `main_shape_simulation.R`, `main_shape_comparison.R`, `main_variance_simulation.R` and `main_shape_application.R`.

`Shape\C_code` is a folder that contains the dynamic libraries and R-wrappers of the C functions.

`Shape\functions.R` is a file that consists of auxiliary R functions.

`Shape\data` is a folder with the file `cetm11659on.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` is a folder that contains simulated distributions of the Gaussian statistic  $\Phi'_T$  for various time series lengths  $T$  and for different grids as well as some deterministic variables for SiZer that are stored in order to decrease running time of `main_shape_comparison.R`. The names of the files with the simulated distributions for Section 5.1 (calculated on the grid  $\mathcal{G}_T$ ) end with `..._constant_type_11.RData` whereas the names of the files with the simulated distributions for Section 5.2, that are based on the abridged

grid  $\mathcal{G}_T^*$ , end with `..._comparison_type_11.RData`. In order to recalculate these distributions, the stored files need to be deleted or removed; the computation is then done automatically while running the main file. The distribution of the Gaussian statistic  $\Phi'_T$  used to calculate the critical values of the test in the application in Section 6 is stored in the file `distr_T_359_testing_constant_type_11.RData`. As before, the distribution of  $\Phi'_T$  can be recalculated by removing or deleting this file before running `main_shape_application.R`.