**A bootstrap application on the standard normal homogeneity test (snht) when there is not highly correlated reference series**

Ceyda Yazıcı*
Middle East Technical University, Department of Statistics, Ankara, Turkey – cyazici@metu.edu.tr

İnci Batmaz
Middle East Technical University, Department of Statistics, Ankara, Turkey – ibatmaz@metu.edu.tr

Ceylan Yozgatlıgil
Middle East Technical University, Department of Statistics, Ankara, Turkey – ceylan@metu.edu.tr

**Abstract**

Climate change possesses many risks for the environment and human beings. The related events such as floods or droughts should be predicted and the precautions should be taken to reduce the damage that may be caused by the extreme events. In order to conduct analyses on climate data, the quality control process should be applied to obtain reliable results. Homogeneity analysis, which is the quality control of the meteorological data, is defined as the detection of non-climatic factors. Standard Normal Homogeneity Test (SNHT) is a widely used homogeneity test applied if only there are highly correlated neighbor series. In this study, a computational method, bootstrapping is applied with SNHT if there is not highly correlated series. The simulation results show that the proposed approach has higher detection rates especially if the break is in the middle of the series.

**Keywords:** time series; homogeneity; moving block bootstrap; F-test.

## 1. Introduction

Climate change has an important effect on the environment and human beings. Extreme weather events such as droughts and floods are the results of climate change and they should be predicted in order to take precautions in advance. Before conducting analyses on related variables, the quality control of the data sets should be conducted. Meteorological variables which include average temperature, minimum temperature, maximum temperature, precipitation, pressure, cloudiness etc. are measured in stations with instruments. Thus, relocations of the stations, urbanization around the stations, changes or any break down in the instruments, or changes or modifications in the calculations have an important effect on the observations. If the changes are originated from any non-climatic measure, this can lead to unreliable results. Otherwise, if these changes are results of climate change, then the series is considered as homogeneous. Thus, inhomogeneity may be caused by abrupt changes, gradual or instant changes or breakpoints and can lead to change in mean, change in variance or change in both of them, or can affect trend. Before conducting any statistical analyses, the quality control of the data should be done to detect and correct the non-climatic effects, if possible. Historical metadata support is the best solution to this problem and essential for evaluating the breaks detected. Unfortunately, most of the data sets do not have accompanying metadata to check the sources of inhomogeneity. The homogeneity tests used in the literature are divided into two groups. The absolute tests need only test series; while the relative tests need highly correlated reference (neighbor) stations. However, the methods used in the literature to detect inhomogeneity have some drawbacks. For instance, the most widely used one Standard Normal Homogeneity Test (SNHT) (Alexandersson, 1986) needs highly correlated reference stations, the nonparametric test such as Kruskal Wallis (Kruskal & Wallis, 1952) do not indicate the exact breakpoint but gives an interval and the number of observations in each block is not determined. Bayesian approaches are graphical approaches which are not effective. The relative tests need homogeneous references, but if there is not prior information, then it is almost impossible to classify the stations.

The most well-known relative homogeneity test is the Standard Normal Homogeneity Test which was proposed by Alexandersson (1986) is a likelihood ratio test. Even though Yazici et al. (2012) show that

SNHT is superior to other tests in terms of detecting inhomogeneity, this test has some drawbacks The test requires reference stations, whose reliability and homogeneity should be validated before the test is conducted. Moreover, similar to other relative tests, it needs close relative so that the correlation between the test station and reference stations must be at least 0.8. If this assumption is not satisfied, then the test station is classified as non-testable (Gokturk et al., 2008).
Since it is the most popular homogeneity test, there are lots of studies which involve this test. For instance, Gokturk et al. (2008) apply SNHT to Turkish monthly precipitation from 267 stations. Hanssen-Bauer & Forland (1994) proposed a four-step approach to define the reference series as homogeneous in SNHT if there is no prior information. Gonzalez-Rouco et al. (2001) used SNHT for the 95 monthly precipitation series of the Southwest Europe and added a one step to the proposed method of Gonzalez – Rouco et al. (2001). Sahin & Cigizoglu (2010) compared the performances of SNHT, Pettitt, von Neumann, Buishand range and the bivariate test developed by Maronna & Yohai (1978) on the 250 meteorological stations of Turkey and concluded that the relative tests are more powerful than the absolute tests. Tayanc et al. (1998) carried out a comparative evaluation by using Kruskal–Wallis and Wald–Wolfz methods to determine the inhomogeneous structure in a temperature series in Turkey. Yozgatlıgil et al. (2011) compared the performances of the nonparametric tests in the homogeneity analysis.
In this study, an application of one of the dependent bootstrap methods is proposed if there is a reference station with correlation smaller than 0.8. Since the data is time series, one of the dependent bootstrap methods called moving block bootstrap (MBB) is applied to construct an empirical distribution of the test statistics and decide whether the test statistics obtained from the original series is insignificant or not by constructing percentile intervals. Moreover, another successful relative test, called F-test is applied to the series to compare the performances of the both methods.
The paper is organized as follows. The aim of the study is stated in Section 1 and the methods used are briefly explained in Section 2. In Section 3, the application is explained and the results are presented in Section 4. In the last section, the conclusions and the further studies are presented.

## 2. Methods

### 2.1. Standard Normal Homogeneity Test (SNHT)

The SNHT is the most widely used test, even though it has a drawback that if there is not any highly correlated reference station, the tested station is left as a non-classified series. The test is based on a likelihood ratio test and aimed to detect the mean shift and locate its position.
The following ratio terms are constructed for precipitation and temperature series, respectively.

$$Q_i = \frac{Y_i}{\left[\sum_{j=1}^{k} v_j X_{ji} \bar{Y} / \bar{X}_j\right] / \sum_{j=1}^{k} v_j}$$

$$Q_i = Y_i - \left\{\sum_{j=1}^{k} v_j \left[X_{ji} - \bar{X}_j + \bar{Y}\right] / \sum_{j=1}^{k} v_j\right\}$$

where Y is the test station, $Y_i$ denotes a specific amount for the year *i*, $X_j$ is the $j^{th}$ reference station and $X_{ji}$ is a specific value from that station. $v_j$ denotes the correlation between that reference station and the test station; and the k indicates the total number of reference stations (Toumenvirta, 2001).
SNHT uses normalized series of ratios for precipitation and differences for temperature, Zi, defined as

$$Z_i = \frac{\left(Q_i - \bar{Q}\right)}{\sigma_Q}.$$

The corresponding hypotheses are given as follows:
$H_0$: Annual values $Y_i$ of the testing variable Y are independently and identically distributed.
$H_1$: A step-wise shift in the mean (a break) is present.

or it can be represented as

$$H_0 : Z_i \in N(0,1) \;\; i \in \{1,\ldots,n\}$$

$$H_1 : \begin{Bmatrix} Z_i \in N(\mu_1,\sigma) \; i \in \{1,\ldots,a\} \\ Z_i \in N(\mu_2,\sigma) \; i \in \{a+1,\ldots,n\} \end{Bmatrix}$$

The test statistic is defined as

$$T^s_{\max} = \underset{1 \le a \le n-1}{Max} \{T^s_a\} = \underset{1 \le a \le n-1}{Max} \{a\bar{z}_1^2 + (n-a)\bar{z}_2^2\}$$

where $\bar{z}_1^2$ and $\bar{z}_2^2$ are the mean values before and after the shift, n is the length of the series. The corresponding mean value of *a* is the most probable break point. The null hypothesis can be rejected, if $T^s_{\max}$ is above the selected significance level.

### 2.2 Modified F-Test

This test is developed by Chow (1960) to determine whether the coefficients of two regression models are equal or not. It is used for testing structural breaks in economics. In the homogeneity analysis, it is possible to use this test as an absolute test; while in this study it is used as a relative test. In one of our studies (Yazici et al., 2012) we proposed to use this test for detecting breakpoints. This test is also capable of detecting the position of the break.

Consider the usual linear regression model;

$$Y_t = X_t\beta + \varepsilon_t, \;\; t = 1,2,...,n;$$

where $X_t$ is the matrix of independent variables, $\varepsilon_t$ 's are the i.i.d. random variables with 0 mean and constant variance. In the homogeneity analysis, the $X_t$ term includes the reference series in the application of relative tests.

The null hypothesis of the test which indicates no structural break is $H_0 : \beta_t = \beta_0$; where the alternative hypothesis indicate that there is a structural change

$$\beta_t = \begin{cases} \beta_A, \; 1 \le t \le K \\ \beta_B, \; K < t < n. \end{cases}$$

Here, *K* indicates the change point. When the test developed, the test is designed to test for a specific point *K*, known in advance. Then, the use of F-test allows to conduct the test or each possible point in the series and give a possible breakpoint.

### 2.3 Moving Block Bootstrap (MBB)

The bootstrap is a computational statistic method which proposes a solution to mathematically intractable problems (Martinez & Martinez, 2002). It is used to estimate standard errors, biases or to construct confidence intervals if assumptions are not satisfied, distribution is not known or there is not a theoretical solution. The bootstrap method is based on treating the data set as a population and drawing samples with replacement from it *B* times. The statistic of interest is calculated from each bootstrap sample and these values obtained *B* times construct the empirical distribution function of that statistic.

There are also bootstrap methods for dependent data, namely subsampling, block bootstrap, moving block bootstrap, sieve bootstrap, the frequency domain bootstrap, etc.

The block bootstrap which is one of the most widely used method has a similar approach to the nonparametric i.i.d. bootstrap. However, in this method, blocks of consecutive observations are taken with replacement instead of single observations in order to reflect the dependency structure. When this method was proposed, nonoverlapping blocks are used with fixed length (*l*). Then, overlapping blocks of length *l* are used, which is defined as the Moving Block Bootstrap (MBB) (Mammen & Nandi, 2012). The method works by selecting $n/l$ blocks and forming the new series of length *n*. In this approach, it is an important issue to decide on the length of the blocks. Politis and White (2004) proposed a method to select the block length automatically based spectral estimation.

MBB works well under weak conditions of the dependency structure, while it does not achieve the accuracy of the bootstrap for the i.i.d. case. Moreover, when it is compared with nonoverlapping blocks, MBB has better higher order properties Bootstrap (Mammen & Nandi, 2012).

## 3. Application

Based on the data in hand, which includes 60 years of monthly values obtained from 244 stations in Turkey, a temperature model is constructed. The best model fitted is the seasonal exponential smoothing method

$$Y_t = \mu_t + S_p(t) + \varepsilon_t$$

where $\mu_t$ is a level parameter representing the mean of the series, $S_p(t)$'s are seasonal parameters and $\varepsilon_t$ is an error term $t = 1,2,\cdots,720; p = seasonality\ period = 12$. The $\varepsilon_t$ term has 0 mean and constant variance. The coefficients estimated from the data are as follows;

$$S_{12}(1) = -7.4, S_{12}(2) = -7.5, S_{12}(3) = -6.1, S_{12}(4) = -3.0, S_{12}(5) = 1.4, S_{12}(6) = 6.2$$

$$S_{12}(7) = 9.0, S_{12}(8) = 9.4, S_{12}(9) = 5.7, S_{12}(10) = 1.8, S_{12}(11) = -2.6 \text{ and } S_{12}(12) = -5.5$$

However, to conduct SNHT, reference stations are needed. To simulate reference stations, similar coefficients are used and two reference series are obtained.

Moreover, a random $\varepsilon_t$ ($\varepsilon_t \sim N_3(\mu = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.6 & 1 & 0.9 \\ 0.8 & 0.9 & 1 \end{bmatrix})$) term is added to all station variables to make the correlation smaller between the test and reference series.

Two inhomogeneity scenarios are considered and applied series. Mean shift may represent the abrupt discontinuity. Since there are 60 yearly aggregates, the 1ºC and 2ºC shifts are applied to the series starting from 5th year (starting from the beginning), 27th year (starting from the middle) and 53th year (starting from the end). The second scenario is the sudden decrease case which represent the change or breakdown in the instrument which measures the meteorological variable. For this case, the yearly aggregate of test station is decreased 1ºC in the 5th year, 27th year and 53th year. SNHT and F-test are applied before creating the break and after creating the artificial break.

After simulating the series, yearly aggregates are obtained and kernel density estimation is conducted based on the Gaussian kernel and the bandwith selection of $0.786 \times IQR \times n^{-1/5}$, where IQR is the interquartile range of the series. The observations associated with the lowest and highest 3 density values are removed from the data to obtain the bootstrap samples. The length selection method of Politis and White (2004) gives the block length of 1 (one) for the yearly aggregates and MBB is applied 250 times. SNHT statistic $T^s_{\max}$ is obtained for each bootstrap sample and the 5% percentile interval (PI) is calculated from the empirical distribution. If the test statistic $T^s_{\max}$ of the original data falls in the PI, then the test station is classified as homogeneous, otherwise it is classified as inhomogeneous. Moreover, the F-test is also applied on the series without conducting bootstrap to compare the performances. The same procedure is repeated after creating the breaks in the data. The whole analysis is repeated 250 times.

## 4. Results

Simulation results are given in the Table 1 and Table 2 for the mean shift and sudden decrease cases respectively. SNHT-BS represents the SNHT test applied with bootstrapping and F-test represents the F-test applied with reference series. $Y_t$ column represents the percentages of inhomogeneity detection before creating the artificial change and $Y_{t,shift}$ column represents the percentages of inhomogeneity detection after creating the artificial change. The percentages obtained from the original series are better for F-test since they are close to Type-I error probability of 5%. However, the detection rates for the SNHT-BS are higher than the F-test. The detection percentages of the SNHT-BS is higher if the break is in the middle of the series in both scenarios. Moreover, the detection percentages increase as the magnitude of the shift increases.

**Table 1. The detection percentages for mean shift**

| Detection rates | 1°C | | | | 2°C | | | |
|---|---|---|---|---|---|---|---|---|
| | $Y_t$ | | $Y_{t,shift}$ | | $Y_t$ | | $Y_{t,shift}$ | |
| | **SNHT-BS** | **F-Test** | **SNHT-BS** | **F-Test** | **SNHT** | **F-Test** | **SNHT-BS** | **F-Test** |
| **At the beginning** | 0.284 | 0.044 | 0.388 | 0.056 | 0.328 | 0.064 | 0.532 | 0.116 |
| **In the middle** | 0.320 | 0.092 | 0.628 | 0.264 | 0.336 | 0.040 | 0.912 | 0.848 |
| **At the end** | 0.296 | 0.052 | 0.444 | 0.108 | 0.300 | 0.060 | 0.696 | 0.340 |

**Table 2. The detection percentages for sudden decrease**

| Detection rates | Sudden Decrease (1°C) | | | |
|---|---|---|---|---|
| | $Y_t$ | | $Y_{t,shift}$ | |
| | **SNHT-BS** | **F-Test** | **SNHT-BS** | **F-Test** |
| **At the beginning** | 0.264 | 0.044 | 0.272 | 0.052 |
| **In the middle** | 0.324 | 0.056 | 0.324 | 0.056 |
| **At the end** | 0.288 | 0.048 | 0.280 | 0.052 |

## 5. Conclusions

The homogeneity analysis is the quality control part of the meteorological studies which should be conducted carefully. The non-climatic effects should be detected and removed if possible to obtain reliable inferences. The most widely used relative homogeneity test, SNHT, needs highly correlated reference series to conduct the test. In this study a computational statistics method; bootstrap for dependent data is applied if there are reference series with correlation less than 0.8. Thus, the non-classified SNHT test stations are tested. The results are compared with another relative test F-test and applied on two inhomogeneity scenarios. The performance of the SNHT-BS method works well especially if the break is in the middle of the series; while it needs to be improved to capture the Type-I probability better.

The improvement of the SNHT-BS is considered to locate the position of the break and capture the Type-I probability better as future studies. Moreover, it will be conducted on more inhomogeneity scenarios.

## Acknowledgements

This research has been supported by Middle East Technical University (METU), Ankara, Turkey, under the project number BAP-01-09-2015-001.

## References

Alexandersson, H. (1986). A homogeneity test applied to precipitation data, *Int. J. Climatol*, **6**, pp. 661–675.

Chow, GC. (1960). Tests of equality between sets of coefficients in two linear regressions, *Econometrica*, **28**(3), pp. 591-605.

Gokturk, O.M., Bozkurt, D., Lutfi, O., & Karaca, M. (2008), Quality control and homogeneity of Turkish precipitation data, *Hydrological Processes*, **22**, pp. 3210 - 3218.

Gonzales-Rouco, J.F., Jimenez, J.L., Quesada, V., Valero, F. (2001). Quality control and homogeneity of precipitation data in the Southwest of Europe, *American Meteorological Society*, **14**, pp. 964-978.

Hanssen–Bauer, I., & Førland, E. (1994). Homogenizing long Norwegian precipitation series, *J. Climate*, **7**, pp. 1001–1013.

Kruskal, W.H, & Wallis, W.A. (1952). Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc*., **47**, 583– 621.

Lahiri, S. N., (2003). *Resampling Methods for Dependent data*. Springer, New York.
Mammen, E., & Nanni, S., (2012). Bootstrap and resampling. *Handbooks of Computational Statistics.* Springer, New York.
Maronna, R., Yohai, V.J. (1978). A bivariate test for detection of systematic change in mean, *J. Amer. Statis. Assoc*., **73**, pp. 640-645.
Martinez, W.L. & Martinez, A.R. (2002). Computational Statistics Handbook with Matlab. New York: Chapman & Hall.
Politis, D.N. & White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, **23** (1), pp. 53-70.
Sahin, S., & Cigizoglu, H.K. (2010). Homogeneity analysis of Turkish meteorological data set. *Hydrol. Process,* **24**, pp. 981–992.
Tayanc, .M, Dalfes, N., Karaca, M., & Yenigun, O. (1998). A comparative assessment of different methods for detecting inhomogeneties in Turkish temperature data set, *Int. J. Climatol*, **18**, pp. 561–578.
Tuomenvirta, H. (2001). Homogeneity adjustments of temperature and precipitation series – Finnish and Nordic data, *Int. J. Climatol*., **21**, pp. 495-506.
Yazici C., Yozgatligil, C., & Batmaz, I. (2012). A simulation study on the performances of homogeneity tests applied in meteorological studies, ICACM: International Conference on Applied and Computational Mathematics Ankara, Turkey. Book of Abstracts, 93.
Yozgatligil, C., Purutcuoglu, V., Yazici, C., & Batmaz, I. (2011). Validity of homogeneity tests for meteorological time series data: A simulation study, Proceedings of the 58th World Statistics Congress (ISI2011).