# COMPARING CHANGES OVER TIME
# FOR TWO PHENOMENA

## Michał Miłek

**Abstract.** Statistical analyses in economics are often based on explaining the phenomena by comparing time series. The purpose of such types of analyses is to find out the similarity and schematic behavior of phenomena which appear in the data. Usual time series are compared with the use of a different similarity measure which, in accordance with the literature, could be divided into four categories. In this article, I propose a method that allows to indicate whether two time series are generated by the same stochastic processes. For this purpose, I analyze a method based on a permutation test. The idea of this test is much simpler than the tests based on theoretical distributions. I also conducted a simulation analysis based on the data generated according to different scenarios, subsequently comparing the results of that analysis.

## 1. Introduction

In economic research practice we are often faced with a situation in which it is necessary to compare time series. This kind of analysis has numerous applications in several fields. In finance, for instance, we want to identify dependences in financial market returns for classifying and grouping stocks; in economics, we tend to draw comparisons between countries and changes in their main macroeconomic indicators; in management we aim at comparing assets for investment and risk management purposes [Caiado, Crato, Peña 2007]. This is also useful in any situation when we want to identify similarities of analyzed phenomena, such as for example in the analysis of geophysical and environmental data or demography. One can find many solutions that can be used to compare time series. Following the categorization introduced by Esling and Agon [2012], the time series distance measures are usually divided into four categories: shape based, edit

---

**Michał Miłek**

   Department of Statistics, University of Economics in Katowice
   michal.milek@edu.uekat.pl

based, features based and structure based. Choosing the right measure heavily depends on the type of data to be analyzed,  as well as the specific properties which may be required. Even though there are no rules that can indicate the proper method of time series comparison, the  literature mentions some clues that can be useful when dealing with such a problem. Methods based on shape are the right choice if the time series are relatively short and can be analyzed in a visual way. The approach seems to be quite different in a situation when we analyze a very specific set of data or we have any prior knowledge about the analyzed phenomenon. When the series are long and very little knowledge on the structure is available, compression based methods seem to be more effective than other methods. The edit distance method was proposed to calculate the similarity between two sequences of strings and is based on the idea of counting the minimal number of edit operations (delete, insert and replace) that are necessary to transform one sequence into the other. The function comparing measure was also analyzed by Cieślak and Jasiński [1979].

In the literature, one can find different proposals of shape based distance measures depending on used distance metrics, for example Manhattan, Minkowski or Euclidean. There are also more sophisticated proposals, such as Dynamic Time Warping (DTW) distance [Berndt, Clifford 1994], LB-Keogh for Dynamic Time Warping [Keogh, Ratanamahatana 2005], Dissim distance [Frentzos, Gratsias, Theodoridis 2007], Short Time Series distance [Moeller-Levet et al. 2003]. Edit Distance for Real Sequences (EDR) [Chen, Özsu, Oria 2005] or Longest Common Subsequence distance (LCSS) [Vlachos, Kollios, Gunopulos 2002] are examples of edit based distance measures.

The purpose of this paper is to indicate whether the observed phenomena change over time in the same way. For this purpose, methods based on a permutation test and four different distance measures were proposed. It is my hypothesis that this solution is effective when comparing changes over time for the two phenomena and allows to identify whether two time series are generated by the same stochastic process. To test the effectiveness of this method, a simulation analysis was conducted based on the data generated according to different scenarios. For each distance measure, the power and size of the test were compared. I decided to start this paper by introducing the theoretical basis of used difference measures and permutation tests. This part is followed by a presentation of the simulation comparison of the used methods and an empirical example of the used method.

## 2. Definitions

The two stochastic processes that I want to compare can be determined by symbols $\{X_t\}_{t=1,2,...} = (X_1, X_2, ...)$ and $\{Y_t\}_{t=1,2,...} = (Y_1, Y_2, ...)$. Using symbols $x_1, x_2, ..., x_k, ..., x_n$ and $y_1, y_2, ..., y_k, ..., y_n$ the elements of the time series can be marked, which are the execution of the processes above. The main purpose is to test if the time series come from the same stochastic processes, which means that $\{X_t\}_{t=1,2,...} = \{Y_t\}_{t=1,2,...}$. Examples of the two compared time series are schematically shown in Figure 1. We can also divide the time series into two parts and compare them to test whether there is a change depending on time. This situation is shown in Figure 2.
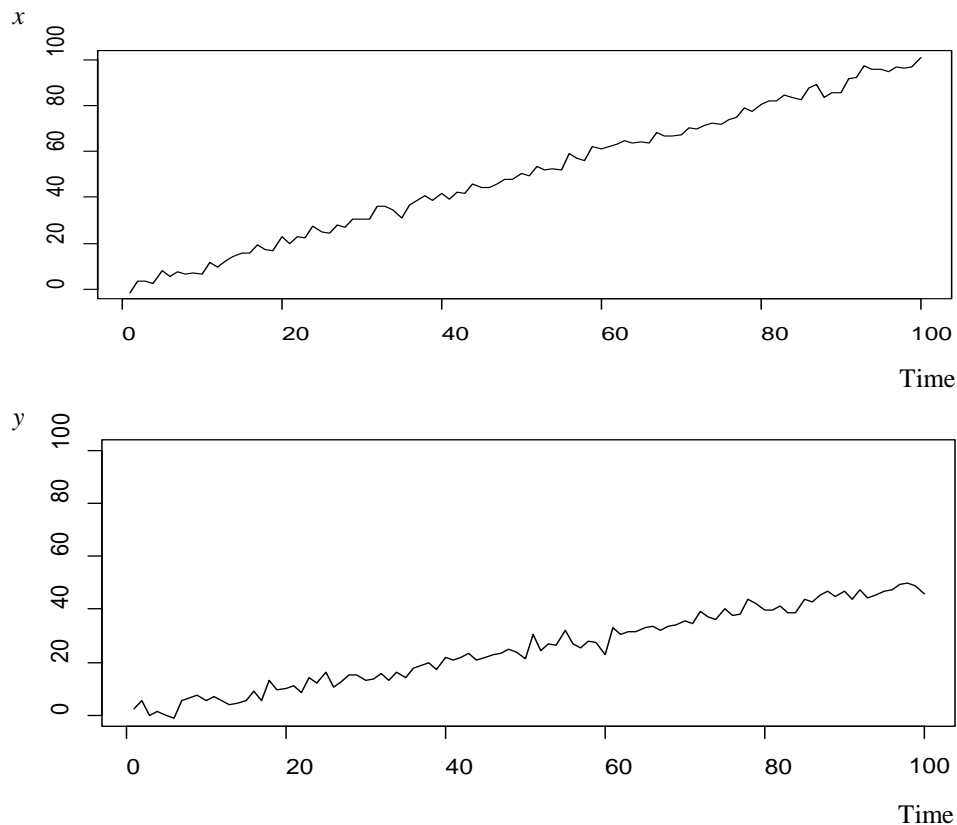


Fig. 1. Example of two time series being compared
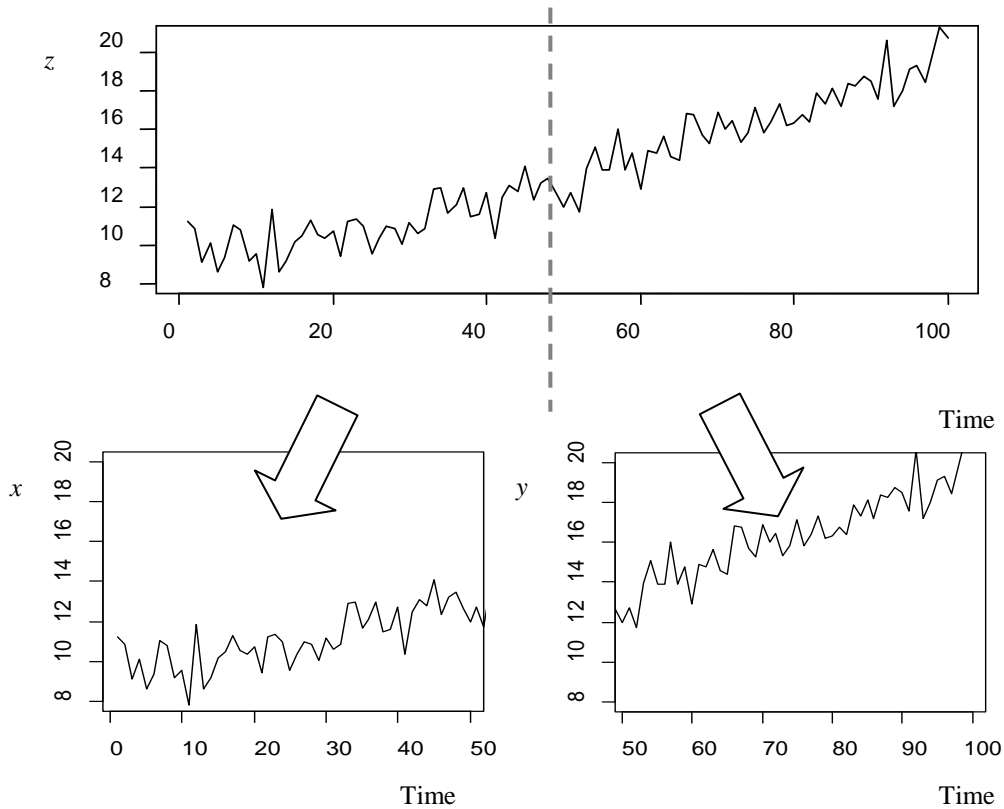
Source: own elaboration.

Fig. 2. Comparing time series in different period of time

Source: own elaboration.

To compare whether these time series come from the same stochastic process, the permutation test was used. Permutation tests are one of the methods of computer-aided simulation and were introduced by R.A. Fisher and E.J.G. Pitman in the 1930s [Kończak 2012]. Permutation tests are normally used for calculating population parameters such as mean, variance and rate structure. Such tests require that the sample was collected from a normally distributed population. For larger samples, limit distributions can be used. In permutation tests, the observed value of the test statistic is compared with the empirical distribution, under the truth of the null hypothesis. The idea of the test is much simpler than the tests based on the normal distribution. The main area of its application is the classic two-sample test [Efron, Tibshirani 1993]. In order to perform the test procedure for permutation tests, these steps should be followed [Good 2005]:

1. Setting the null hypothesis and the alternative.

2. Determine the form of the test statistics ($T$).

3. Calculate the value of the test statistics ($T_0$) on the basis of empirical data.

4. Evaluation of the distribution of the test statistics based on simulations assuming the veracity of $H_0$ ($T_1$, $T_2$, …, $T_N$, where $N > 1000$).

5. Decide on the basis of the resulting distribution of the test statistics.

A decision is made on the basis of the value of *ASL* – Achieving Significance Level (*p*-value equivalent in standard statistical tests). *ASL* is expressed as follows:

$$ASL = P\left(\hat{\theta} \geq \hat{\theta}_0\right). \tag{1}$$

This value can be estimated based on empirical data using the following formula:

$$ASL \approx \frac{card\{i : \hat{\theta}_i \geq \hat{\theta}_0)}{N}. \tag{2}$$

The tests based on some distance measures used in this paper require left-handed critical area. In such a situation, in order to compute the *ASL* value, we can use formulas as follows:

$$ASL = P\left(\hat{\theta} \leq \hat{\theta}_0\right), \tag{3}$$

$$ASL \approx \frac{card\{i : \hat{\theta}_i \leq \hat{\theta}_0)}{N}. \tag{4}$$

This means that only the appropriate changes in the inequalities are needed. If $ASL < \alpha$ it is decided to reject $H_0$ in favor of $H_1$. Otherwise, there is no basis to reject this hypothesis.

In the part that follows, a permutation test was used to decide if there is a significant difference in changes of tempo in the two time series. The form of the test statistic for permutation tests depends only on the investigator. In order to compare the time series, the following methods were used: LB-Keogh for Dynamic Time Warping, the Dissim Distance approximation, the Euclidean distance measure and the Manhattan distance measure. All the solutions are in the group of shape distance measures. The formulas of the test statistics were the difference measures mentioned above, that is:

- LB-Keogh for Dynamic Time Warping

$$T^{(1)} = d_{LB\_Keogh}(X,Y) = \begin{cases} (Y_i - U_i)^2 & \text{if} & Y_i > U_i \\ (Y_i - L_i)^2 & \text{if} & Y_i < L_i \\ 0 & \text{otherwise} \end{cases}, \qquad (5)$$

where: $r$ – range of warping and the upper and lower envelope series are defined as follows:

$$U_i = \max(X_i - r, X_i + r), \ L_i = \min(X_i - r, X_i + r).$$

- Dissim Distance approximation:

$$T^{(2)} = d_{Dissim\_app}(X,Y) = \sum_{i=0}^{K-1} \big( D_{X,Y}(t_i) + D_{X,Y}(t_i) \big) \cdot (t_{i+1} - t_i). \qquad (6)$$

- Manhattan distance measure:

$$T^{(3)} = d_{Manhattan}(X,Y) = \sum_{i=0}^{N-1} |x_i - y_i|. \qquad (7)$$

- Euclidean distance measure:

$$T^{(4)} = d_{Euclidean}(X,Y) = \sqrt{\sum_{i=0}^{N-1} (x_i - y_i)^2}. \qquad (8)$$

The test procedure of a permutation test is as follows:

1. Establishment of the level of significance $\alpha$.

2. Calculate the $T_0$ value of statistic $T$ based on simulated data.

3. Execution of the time series permutations of $N$ times, then calculating the value of the test statistics.

4. On the basis of the empirical distribution of the test statistics $T$, the *ASL* value is calculated.

If $ASL < \alpha$, then the hypothesis $H_0$ is rejected, otherwise there is no basis to reject $H_0$.

As the number of repetitions of permutations assumed $N = 1000$.

### 3. Comparison of the proposed methods and an empirical example

I carried out a computer simulation using different scenarios. The first scenario assumes that both the time series: $x_1, x_2, ..., x_k, ..., x_n$ and $y_1, y_2, ..., y_k, ..., y_n$ were generated by the same stochastic process, which

means that $\{X_t\}_{t=1,2,...} = \{Y_t\}_{t=1,2,...}$, according to the formula (9). The proposed model is only one of the suggestions, the scope of this test is much wider, as shown in the empirical part.

$$x = a_1 t + e$$
$$y = a_2 t + e \tag{9}$$

where $e = N(0, 1)$.

The next scenario assumes that $\{X_t\}_{t=1,2,...}$ and $\{Y_t\}_{t=1,2,...}$ are not the same stochastic processes. In the simulation, the following levels of parameters were analyzed: $a_1 = 1$ and $a_2$ were on different levels: $a_2 = \{1, 0.99, 0.97, 0.95, 0.90, 0.80\}$.
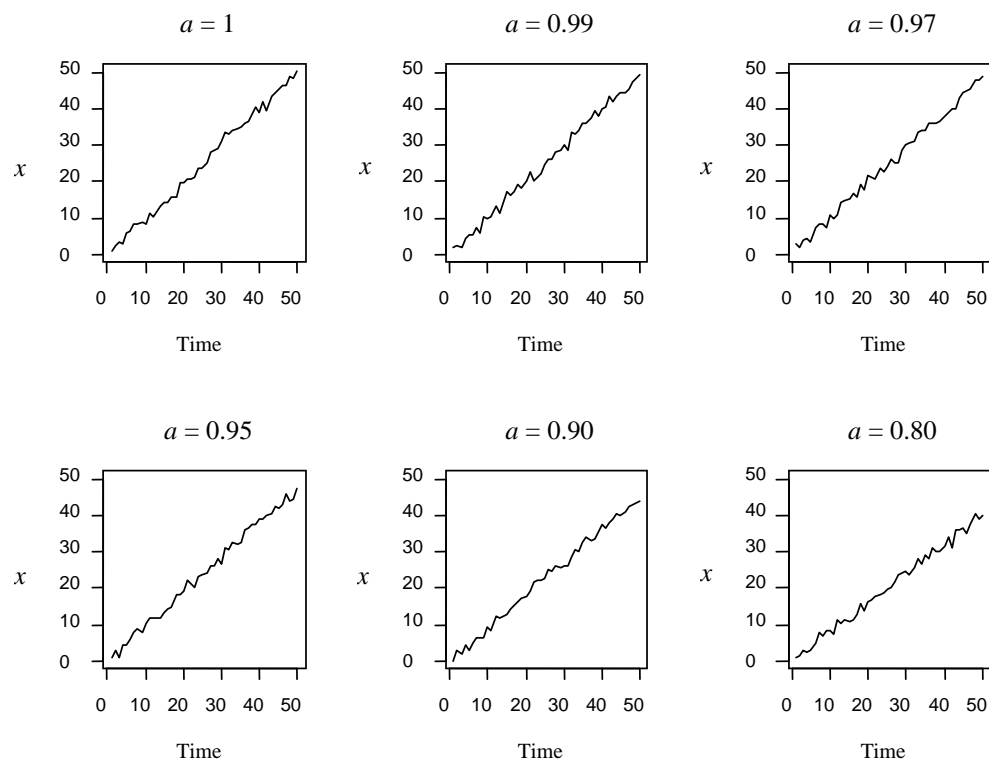


Fig. 3. Example of time series considered in analyses

Source: own elaboration.

Table 1. Size and power of tests

| Test statistic | Size of test | Power of test | | | | |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| | $a_2 = 1.00$ | $a_2 = 0.99$ | $a_2 = 0.97$ | $a_2 = 0.95$ | $a_2 = 0.90$ | $a_2 = 0.80$ |
| $T^{(1)}$ | 0.054 | 0.950 | 0.959 | 0.962 | 0.991 | 0.998 |
| $T^{(2)}$ | 0.042 | 0.966 | 0.979 | 0.99 | 0.999 | 1.000 |
| $T^{(3)}$ | 0.049 | 0.958 | 0.977 | 0.995 | 1.000 | 1.000 |
| $T^{(4)}$ | 0.054 | 0.968 | 0.978 | 0.992 | 0.999 | 1.000 |

Source: own elaboration.

At each iteration of the simulation, two time series were generated according to the rules explained above. For each distance measure, the power and size of test were compared. Examples of the time series under consideration are shown in Figure 3. Table 1 contains the simulation results which are also presented in Figure 4.
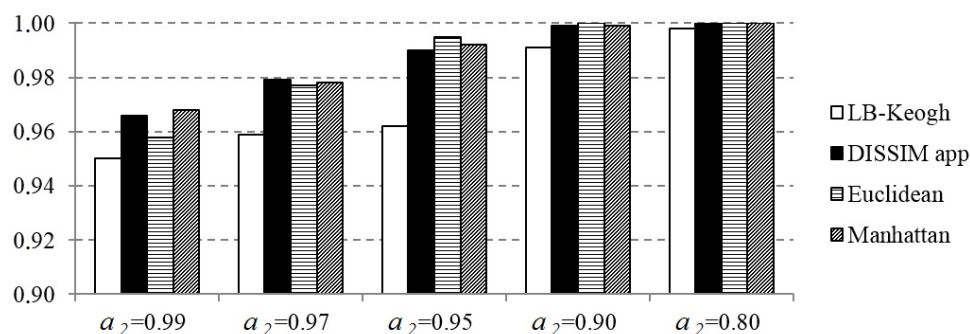


Fig. 4. Power of tests

Source: own elaboration.

Table 2. Results of similarity testing in empirical examples

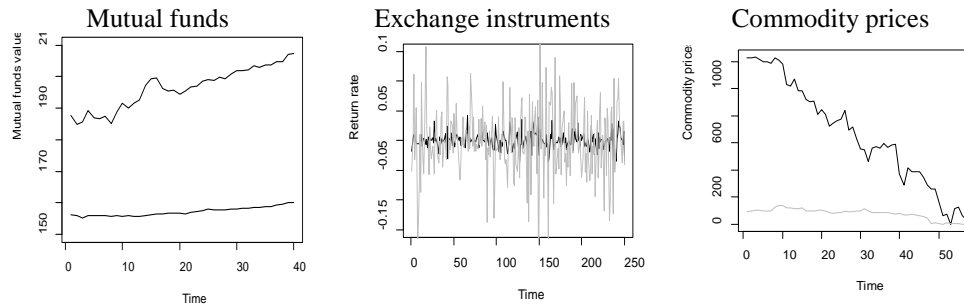| Test statistic | ASL | | |
|:---:|:---:|:---:|:---:|
| | Mutual funds | Exchange instruments | Commodity prices |
| $T^{(1)}$ | 0.47 | 0.00 | 1.00 |
| $T^{(2)}$ | 0.53 | 0.10 | 0.49 |
| $T^{(3)}$ | 0.53 | 0.01 | 0.48 |
| $T^{(4)}$ | 0.54 | 0.35 | 0.45 |

Source: own elaboration.

Fig. 5. Data analyzed in empirical example

Source: own elaboration.

The proposed tests were used to compare three different data sets: the value of two mutual funds of one Polish bank, the rate of return of the two exchange instruments and the value of copper and gold (the levels of both instruments have been artificially equalized). The analyzed data sets are presented in Figure 5 and the testing results in Table 2. According to the results, it can be concluded that only one of the analyzed phenomena changes over time in the same way. In the case of the two other pairs of data it is clear that there is a significant difference in the changes.

## 4. Conclusion

The permutation tests were used to compare changes over time for the two phenomena. These tests do not meet the assumption as when using the parametric tests. For this purpose, the Dissim Distance approximation, the Euclidean distance measure and the Manhattan distance measure were used. The computation of the power and size of permutation test allowed us to examine the effectiveness of the method. Based on the results above coming from the simulation analysis, it can be concluded that permutation tests are possible to be used in order to decide if there is a significant difference in the shape of time series. There were different results according to the method used. The best results were obtained by using the Manhattan distance measure. The least accurate results were given by the LB-Koegh method. Simulation studies have shown that the proposed method is effective for time series generated by linear process. Further studies can be extended to comparing time series generated from other processes as shown in the empirical part.

# References

Berndt D.J., Clifford J. (1994). *Using dynamic time warping to find patterns in time series*. In KDD Workshop. Vol. 10. No. 16. Pp. 359-370.

Caiado J., Crato N., Peña D. (2007). *Comparison of time series with unequal length*. Munich Personal RePEc Archive.

Cieślak M., Jasiński R. (1979). *Miara podobieństwa funkcji*. Przegląd Statystyczny 3(4).

Chen L., Özsu M.T., Oria V. (2005). *Robust and Fast Similarity Search for Moving Object Trajectories*. In: *Proceedings of the 2005 ACM SIGMOD International Conference On Management of Data*. ACM. Pp. 491-502.

Domański C., Pekasiewicz D., Baszczyńska A., Witaszczyk A. (2014). *Testy statystyczne w procesie podejmowania decyzji*. Wydawnictwo Uniwersytetu Łódzkiego. Łódź.

Efron B., Tibshirani R.J. (1994). *An Introduction to the Bootstrap*. CRC Press.

Esling P., Agon C. (2012). *Time-series data mining*. ACM Computing Surveys (CSUR) 45(1). Article No. 12.

Frentzos E., Gratsias K., Theodoridis Y. (2007). *Index-based Most Similar Trajectory Search*. In: *Proceedings of International Conference on Data Engineering*. Pp. 816-825.

Good P. (2013). *Permutation Tests: a Practical Guide to Resampling Methods for testing Hypotheses*. Springer Science & Business Media.

Keogh E., Ratanamahatana C.A. (2005). *Exact indexing of dynamic time warping*. Knowledge and Information Systems 7(3). Pp. 358-386.

Kończak G. (2012). *Wprowadzenie do symulacji komputerowych*. Wydawnictwo Uniwersytetu Ekonomicznego.

Möller-Levet C.S., Klawonn F., Cho K.H., Wolkenhauer O. (2003). *Fuzzy Clustering of Short Time-Series and Unevenly Distributed Sampling Points*. In: *Advances in Intelligent Data Analysis V*. Springer. Berlin Heidelberg. Pp. 330-340.

Vlachos M., Kollios G., Gunopulos D. (2002). *Discovering Similar Multidimensional Trajectories*. In: *Proceedings of 18th International Conference on Data Engineering*. Pp. 673-684.