Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

MASTER'S THESIS

Transition Periods and Long Memory Property

Author: **Bc. Jan März**

Supervisor: **Mgr. Lukáš Vácha, Ph.D.**

Academic Year: **2014/2015**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, June 15, 2015

Signature

Acknowledgments

The author would like to thank to Mgr. Lukáš Vácha, PhD. for all the valuable ideas and comments in all phases of the work on this thesis.

Abstract

This thesis examines the relationship between the distribution of structural breaks within a data sample and the estimated parameter of long memory. We use Monte Carlo simulations to generate data from processes with specific values of parameters. Subsequently we join the data with various shifts to mean and examine how the estimates of the parameters vary from their true values. We have discovered that the overestimate of the long memory parameter is higher when the breaks are clustered together. It further increases when the signs of the shifts are positively correlated within the clusters while negative correlation reduces the bias. Our findings enable the improvement of robustness of estimators against the presence structural breaks.

JEL Classification	C22, C50, C10
Keywords	fractional integration, long memory, structural break, clustering

Author's e-mail	jan.marz@seznam.cz
Supervisor's e-mail	vachal@utia.cas.cz

Abstrakt

V této diplomové práci zkoumáme vztah mezi výskytem strukturálních změn během daného časového období a odhadem parametru dlouhé paměti. Na základě procesů s přesně danými parametry simulujeme data a uměle upravujeme jejich střední hodnotu. Následně analyzujeme, jak se liší odhad parametru frakční integrace od jeho skutečné hodnoty. Zjistili jsme, že tento odhad je více nadhodnocený, pokud jsou strukturální změny seskupeny v čase. Tato chyba je ještě větší v případě, kdy jsou znaménka těchto strukturálních změn kladně korelovaná. Naopak záporná korelace nadhodnocení snižuje. Naše zjištění umožňují posílení odolnosti odhadu parametru frakční integrace vůči přítomnosti strukturálních změn.

Klasifikace	C22, C50, C10
Klíčová slova	frakční integrace, dlouhá paměť dat, strukturální změny

E-mail autora	jan.marz@seznam.cz
E-mail vedoucího práce	vachal@utia.cas.cz

Contents

List of Tables

List of Figures

Master's Thesis Proposal

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Author:	**Bc. Jan März**	Supervisor:	Mgr. Lukáš Vácha Ph.D.
E-mail:	jan.marz@seznam.cz	E-mail:	vachal@utia.cas.cz
Phone:	+420 602 718 525	Phone:	+420 602 161 710
Specialization:	FFMaB	Defense Planned:	June 2015

Proposed Topic:

Transition periods and long memory property

Motivation:

In my diploma thesis I intend to examine and discuss the topic of long memory property, which can be generated either by an I(d) process present in the data or by structural breaks. Granger & Hyung (2004) show in their paper that these two phenomenons are easy to be mistakenly interchanged when trying to identify the data generating process and the cause of the long memory, which fact leads to decreased forecasting performance. To demonstrate the complication, authors employ a simple occasional break model and test the simulated data for an I(d) process. In the theis, I intend to introduce a group of more sophisticated models for structural breaks, examining outcomes of tests for long memory processes and their dependence on parameters of the models.

A question of highest interest to me is the consequence of structural breaks' clustering, i.e. of occurrence of certain transition periods, and its impact on the long memory properties estimated in the simulated data. The two possibilities as seen prior to doing the research are:
1) When the breaks happen in 'groups' (clustered together), their effect corresponds to one of a proportionally magnified break (or diminished, based on the direction of particular changes), as on bigger scale they can be perceived as one turbulence. However as we are speaking of clustering of breaks, i.e. grouping them together in time, the aggregate change remains unaffected and so do the overall properties of data including the long memory. Such scenario would be in line with the research of Granger & Hyung (2004), and the simple occasional break model with shifts in mean would be an efficient generalization.
2) From the point of view of spectral analysis, different degree of clustering introduces differently low 'frequency' to the data (meant as a consequence for the long memory estimation). In that case I would attempt to identify some basic pattern and find the rationale behind it.

There are more parameters to be adjusted, such as the degree of 'volatility clustering' (or in general probability of occurrence dependent on size of previous break), the lag of probability dependence or shifts in volatility. In all cases there will be danger of introducing a non-linearity to the model or breaking some other assumption which will have to be considered.

Dependent on the extent and difficulty of the above described research I will consider employing the STAR (Smooth Transition Autoregressive) model to define the progress of the transition period, getting closer to real world situations when not all changes happen in form of sudden breaks.

Furthermore, having the simulated data from the various data generating processes there is an option of extending the research to test performance of the method introduced by Wang, Bauwens & Hsiao (2012) on the data with clustered breaks. Their AR-based approach is suited for forecasting an ARFIMA process with unknown dates of shifts to mean and to long memory parameter. I will possibly test the performance of this approach on the processes with changes distributed based on rules specified in the above described part of my thesis.
My thesis will contribute to the current state of research by elaborating on very up-to-date and frequently cited working papers and by extending and detailing their conclusions.

Hypotheses:

1. Hypothesis #1: The estimated long memory parameter depends on the degree of structural breaks clustering.
2. Hypothesis #2: The estimated long memory parameter depends on the lag of structural breaks clustering.
3. Hypothesis #3: Clustering of same-sign shifts to mean has different effects on the estimated long memory parameter than sign-varying fluctuations.

Methodology:

Substantial part of my thesis will be based on Monte Carlo simulations. First step will be defining the structural breaks distribution and the data generating processes according to the theoretical assumptions.
For testing of the first hypothesis is essential the probability of occurrence of a break conditional on past breaks. If the long memory property present in the generated data does not change significantly with this parameter the first hypothesis will be rejected. Extension to this testing can be conditioning the probability and magnitude of a shock on the size of past breaks.
The second hypothesis will be tested using modifications to the 'persistence' of shocks, i.e. how long to the future is the probability of another break affected by them. This hypothesis will be rejected if there is no significant change to the estimated long memory with varying lag in the breaks' distribution process.
Testing of the third hypothesis will be done by letting the sign of a shock affect the direction of consequent shock(s). If the estimated long memory parameter does not change significantly with the probability of the same-sign shock this hypothesis will be rejected. It can be tested both in conjunction with the first hypothesis (combining the effect of probability of occurrence and of probability of same-sign effect) and separately.
After the structural breaks generating and data generating processes are defined in each step, I will simulate data samples and estimate the (time varying) long memory property. Based on the results, I will assess correctness of the hypotheses.

Expected Contribution:

In my thesis I will extend the conclusion of a frequently cited paper and provide more detail for one of its major results. The original paper served as an incentive to develop estimators of long memory resilient to structural breaks. My research can help create their alternative versions tailored for cases when there is more information known or assumed about the distribution of structural breaks in the sample, thereby increasing the quality of forecasts.

Outline:

First part of the thesis will be a brief introduction to the problem and overview of the current state of research. In the second part the data generating processes will be defined and the models for structural breaks determined. In the third part I will analyze the results. In part four I will test the performance of Wang, Bauwens & Hsiao estimator on the simulated data. The final part will be conclusion and summary of most significant results.

Core Bibliography:

GRANGER, Clive WJ; HYUNG, Namwon. Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of empirical finance*, 2004, 11.3: 399-421.

WANG, Cindy Shin-Huei; BAUWENS, Luc; HSIAO, Cheng. Forecasting a long memory process subject to structural breaks. *Journal of Econometrics*, 2013, 177.2: 171-184.

SHIMOTSU, Katsumi, et al. Exact local Whittle estimation of fractional integration. *The Annals of Statistics*, 2005, 33.4: 1890-1933.

DIEBOLD, Francis X.; INOUE, Atsushi. Long memory and regime switching. *Journal of economics,* 2001, 105.1: 131-159.

BAI, Jushan; PERRON, Pierre. Computation and analysis of multiple structural change models. *Journal of applied econometrics,* 2003, 18.1: 1-22.

Author **Supervisor**

1 Introduction

Distinguishing between the evidence of the long memory induced by a fractionally integrated process I(d) and the one brought by the occurrence of structural breaks can be problematic. Mistakenly interchanging these two causes can lead to biased forecasts (Granger and Hyung, 2004). Even separately, structural breaks within the sample can cause a bias in the forecast when not treated properly (Clements and Hendry, 1998). The same goes for the long memory parameter in case of its incorrect estimation (Reisen and Abram, 1998; Reisen and Lopes, 1999). Granger and Hyung (2004) tested the forecasting efficiency of an occasional breaks model against an I(d) model and based on the results they suggested that a plausible replacement for the former model should be chosen. Wang, Bauwens and Hsiao (2013) show an alternative way, approximating an ARFIMA(p,d,q) process with breaks by an AR(k) process. Their approach shows better forecasting performance than the simple ARFIMA-based methods. Using their approximation also allows to avoid identifying the breaks and estimating the parameters of the original model. The cost of the suggested method is the necessity to determine the degree k correctly. Another method of accounting for structural breaks is suggested by Xu and Perron (2014) who extend the work of Lu and Perron (2010) on the random level shift model by adding two modifications to it – the time varying probability of shifts and a mean reverting mechanism. Unlike the ARFIMA models, their method is able to predict also the structural breaks, even though with some limitations. A relatively extensive research has been done on the topic of structural breaks' identification (e.g. Busetti, 2002; Bai and Perron, 2003; Perron and Zhu, 2005). In contrast to that, the understanding of how the estimate of the long memory parameter behaves given specific breaks' distribution models is still very limited The research done concerns mostly the direct influence of structural change on the long memory – switching between regimes with different true value of the parameter d (Hidalgo and Robinson, 1996). Beine and Laurent (2000) provide evidence that the presence of structural breaks biases the estimates of the long memory without going further into the detail into how this relationship works. The same holds for Krämer and Sibberstein (2000) who focus on the identification of structural breaks under long memory.

Wang, Bauwens and Hsiao (2013) and Xu and Perron (2014) suggest replacements of ARFIMA based models with various advantages and disadvantages. Their models build only on the fact that the evidence of the two sources of long memory is interchangeable. Instead of going in that direction we examine how the estimate of the long memory parameter d of an I(d) process responds to different timing of structural breaks within the data sample. We believe that a detailed understanding of the relationship is necessary prior to examining how to correct the bias. For that purpose we analyze the effect of clustering of breaks based on Monte Carlo simulations. The clustering can have various properties, most importantly the degree, which quantifies its intensity, and the length, which controls the extent of the clusters. Our intention is to evaluate what impact changing of these properties in various ways exactly has on the estimated parameter d. By this we want to enable developing tools for estimating the long memory parameter that are resistant to structural breaks. If the estimate of the parameter is precise enough and the structural breaks are correctly identified the approximation of the ARIMA(p,d,q) process by an AR(k) process can be avoided.

The thesis is structured as follows: Chapter 2 describes the general methodology of our simulations, explains how we measure the degree of clustering and how we estimate the long memory parameter. Chapter 3 details what data generating processes we employ for the testing of our three hypotheses and what expectations we have about their outcome. Chapter 4 describes the results of the simulations and the estimates of the long memory parameter and provides the comparison of results to our expectations. Chapter 5 summarizes our findings.

2 Methodology

2.1 The Long Memory and Its Estimation

The presence of low frequency fluctuations in economic data has been first indentified by Granger (1966). He has noted in his research examining the shapes of spectra of the economic data that even after the removal of the trend the fluctuations with very long wave length (exceeding the sample length) are identified as especially important. In regards of a sample from the process with typical shape of spectrum, Granger says:

"The most obvious property of such a sample would probably be a visual long-term fluctuation which would not be in any way periodic. The fluctuation need not be visible at all moments of time and would, of course, be somewhat blurred by the high frequency and seasonal components." (Granger, 1966)

The high frequency components can be represented by the AR and MA elements in our models. Based on the findings of Granger and Hyung (2004) we know that the long term fluctuations can be simulated by adding artificial shifts to mean to the generated data. Therefore we need to decide about what specific distributions of the breaks within the sample we should use. Because we want to examine the connection between the parameters of the distribution and the estimate of the long memory we need to determine what these parameters are.

Since Granger and Hyung (2004) have already researched the effects of evenly distributed breaks (those from an occasional breaks model) we want to focus on uneven distributions. That leads us to the concept of clustering of breaks. We identify two basic properties of the clustering: its degree and its extent. The former property describes the density of breaks within clusters and can be determined by the additional probability of break's occurrence after another break. The latter property expresses how many subsequent periods are influenced by a break's occurrence. The connection between the estimated long memory parameter and the degree and extent of clustering will be the focus of the first and second hypotheses, respectively. Additionally we theorize that there is also a connection between the correlation of signs of individual shifts to mean and the measured long memory, for which reason we add the third hypothesis. All hypotheses are stated formally in the respective parts of Chapter 3.

A useful formal definition of long memory can be found in Beran et al. (2013):

"Let X_t *be a second-order stationary process with autocovariance function* $\gamma_X(k)$ $(k \epsilon Z)$ *and spectral density*

$$f_X(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma_X(k)\, exp(-ik\lambda), \qquad (\lambda \epsilon [-\pi; \pi])$$

Then X_t *is said to exhibit long-range dependence if*

$$f_X(\lambda) = L_f(\lambda)|\lambda|^{-2d}$$

Where $L_f(\lambda) \geq 0$ *is a symmetric function that is slowly varying at zero, and* $d \,\epsilon\, (0, \frac{1}{2})$*."* (Beran et al., 2013)

It defines the long memory based on the spectral density of the process. The link to the frequency domain is important; however the following definition which avoids it seems to be more intuitive:

"Let X_t *be a stationary process for which the following holds. There exists a number* $\alpha \,\epsilon\, (0,1)$ *and a constant* $c_\rho > 0$ *such that*

$$\lim_{k \to \infty} \rho(k) / \left[c_\rho k^{-\alpha}\right] = 1.$$

Then X_t *is called a stationary process with long memory or long-range dependence or strong dependence, or a stationary process with slowly decaying or long-range correlations."* (Beran, 1994)

In this case the parameter d corresponds to 0.5 – $\alpha/2$. It is easy to see the idea of long memory from this second definition. Simply put, with increasing time distance between two observations the autocovariance decays with the distance to the power of (1 – 2d).

There are several methods of estimating the long memory parameter d. First is the log-periodogram regression (Robinson, 1995). Second is the local Whittle estimator as defined by Beran (1994). These two are not used in this thesis. We decide to use a third possibility, the exact local Whittle estimator, an extension derived by Shimotsu et al. (2005), which is defined as follows:

$$\hat{d} = \underset{d \in [\Delta_1, \Delta_2]}{\arg\min}\, R(d)$$

$$R(d) = \log \hat{G}(d) - 2d \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j\,, \qquad \hat{G}(d) = \frac{1}{m} \sum_{j=1}^{m} I_{\Delta^d X}(\lambda_j)$$

Where $-\infty < \Delta_1 < \Delta_2 < \infty$ *are the lower and upper bound of the admissible values of d and* $I_{\Delta^d X}(\lambda_j)$ *is the periodogram of*

$$\Delta^d X_t = (1 - L)^d X_t.$$

Expressed by words, $\hat{d}$ minimizes the difference between logarithm of average value of periodogram of d-differentiated sequence, and 2d times average of logarithm of frequency. Intuitive explanation is that we are finding such d that the periodogram of the d-differentiated sequence corresponds to that of the real data. In the definition m stands for the frequency at which the periodogram is truncated while we are expressing it as a portion of the sample length. The exact version of the local Whittle estimator was derived for the purpose of estimating the long memory parameter of non-stationary processes (d $\geq$ 1/2) for which it possesses several advantages compared to its original version. The estimator is proven to have the same N(0,1/4) limit distribution – under two simple conditions – for all values of the parameter d, unlike the original version.

The value of the parameter m, indicating the truncation of frequencies which are used in the estimation of the long memory parameter, needs to be set. The parameter has values on the interval $(0,1]$, where the value of 1 implies truncation at frequency corresponding to one half of the sample length. In case that the estimated process is not pure ARFIMA(0,d,0)[1] the estimate of parameter d can be contaminated by the short memory part of the process realized in higher frequencies. We want to avoid this by using the truncation. The cost of not using all of the frequencies for the estimation has the form of a trade-off between bias and variance of the estimates. See the Figure 1 for representation of this trade-off based on an example of 100 repetitions of an ARFIMA(α_1 = 0.1, d = 0.3) process of length t = 1000, where α_1 is an AR coefficient. Given the results of this simulation we choose to work with parameter m = 0.4 which corresponds to truncation at one fifth of sample length. Below this point the increase in variance of the estimator gets steeper and more than half of the bias is mitigated compared to the non-truncated version.

2.2 General approach to simulations

A starting point for all of the simulated data is in all cases some common data generating process – for example ARMA with low degree of both AR and MA – and it is always stated which initial process is used as the core. To check for the robustness of results several initial processes are used while leaving other parameters unchanged.

[1] For the definition of an ARFIMA process see for example Granger and Joyeux (1980) or Hosking (1981)

The aim is to create a process in which a “structural change” occurs at given points in time. The simplest example of such change is a shift in mean of the process. Without loss of generality we can have the original process with zero mean and then add another time series containing the shifted mean. This second series is basically a random walk with innovations entering only in some periods, based on whether a break occurs or not. The process of cumulated shifts has the following formula:

$$Shift_t = Shift_{t-1} + \varepsilon_t * Break_t, \qquad \varepsilon_t \sim \Phi(0, \sigma_\varepsilon^2), \tag{1}$$

Throughout the whole thesis, Φ stands for normal distribution. It would be hard to evaluate the results from testing of the hypotheses only based on comparison of two situations, with and without the added shift to mean. For this reason we always introduce the additional probability gradually. We take the target additional probability as a maximum value for the instance of the experiment. We split the simulation into steps and linearly increase the additional probability from zero value up to this maximum. With such setting we can examine how the degree of clustering of the breaks in the simulated data evolves and how the estimate of the long memory parameter changes. Our approach increases the robustness of the estimate of relationship between the clustering and long memory and it also allows us to draw more detailed conclusions. We artificially set the number of steps to 20. The concept of steps is important mainly for the part of this thesis dedicated to the results of testing. Every estimate is presented on a scale from 1 to 20 starting in the situation without tampered distribution of breaks and ending with vector of probabilities equal to their maximum value for the particular instance of the experiment. Each step of the experiment is repeated 200 times.

The vector of breaks is created in the following way: first a vector v of values from uniform distribution on the interval $(0,1)$ is generated. Then each value is compared to the respective probability

$$v_t \lesseqgtr p_0 + p_1 * Break_{t-1}, \qquad t = 1 \dots n. \tag{2}$$

The parameter p_0 stands for the independent probability of break occurrence and its value is discussed later on. The parameter p_1 is the additional probability of break. If v_t is smaller than the right hand side of the Equation 2 then a break occurs at the period t.

2.3 Measure of Clustering

In order to reject or not reject all of the hypotheses it is necessary to be able to quantify the results. We are trying to explain the estimated long memory parameter by the ambiguous "degree of clustering" while the experiment as described above yields only the vectors of breaks.

Structural breaks are more clustered in time in one situation than in another one when there is lower number of observations between them. It is important to realize that the overall time span in which all of the breaks happen – the time period of the last one minus the time period of the first one – is not the primary value we care about. It is only a sum of all of the intervals between individual breaks.

An example can be thought as follows: imagine only four breaks happening during the whole sample. In the first situation, the breaks are evenly distributed over the whole sample. In the second situation two of them occur in the first two periods and two of them in the last two periods. The overall time span is in both cases the whole sample and the mean distance between the breaks is also equal in the two cases. Our measure has to be able to differentiate between such variants since the clustering is clearly more intensive in the second situation. Therefore we suggest a relative measure which compares the realized distribution to two extremes – the no clustering situation and the perfect clustering situation.

The latter one is quite easy to define – the breaks are perfectly clustered together if they all occur in an uninterrupted sequence starting at any point of the time sample. For the opposite situation we say that there is absolutely no clustering present in the data when the breaks occur as in the situation described above – evenly distributed over the whole sample – so the mean distance in time between two breaks is the highest possible, equal to

$$Maximum\ mean\ breaks'\ distance = \frac{Number\ of\ time\ periods - 1}{Number\ of\ breaks - 1} \tag{3}$$

since we suppose that in case of perfect clustering the whole time sample is used and first and last time period always contain a break. For precision in the calculation of maximum mean breaks' distance the realized number of breaks is used, not the expected number, since we are comparing it to the realized distance. It is also worth mentioning that by distance between two time periods we refer to the number of gaps between them, not to the number of time periods between them (which means that the sum of distances between breaks in a sample of n time periods is (n – 1)).

Because in our perfect-clustering situation the breaks are evenly distributed over the whole sample we need to evaluate the realized situation also on the scale of the whole sample. That is, even though there does not need to be a break in the first or in the last period, we need to include the distance between first period and first break as well as the distance between the last break and the last period in the evaluation. If it was not done so then the two following situations would not be comparable to the same maximum mean distance, causing bias to the relative measure: when all breaks are evenly distributed over the whole sample and when they are distributed evenly only over a part of it.

For the definition of the no-clustering situation we are using the mean distance between breaks. However, mean is not an appropriate measure for the realized distance. Since we are always splitting the whole time sample for reasons described in the previous paragraph, the sum of distances remains the same in all cases and so does the mean. However with stronger clustering the number of shorter distances between breaks increases relatively to the number of longer breaks. Thus the median time distance between breaks seems to be an appropriate measure.

The disadvantage of using the median is that it is not sensitive to changes in breaks distribution which do not cause a distance to cross the median value. In other words, if a short distance (lower than the median) becomes even smaller or a high distance (higher than the median) becomes higher, the median value remains unchanged. This could seem to be an unfortunate flaw since we would like to capture the effect of those small changes as well. However on large sample (many time periods) with higher parameter p_1 the number of breaks with distance between them equal to one increases enough for the median to capture it. Only the change in median induced by increased number of distances of length one is of essence for testing of the first hypothesis as we are looking only one time period back. For the second and third hypothesis all breaks which occur in the past periods up to the maximum number of lags are essential.

Based on the above discussion, we define the measure of clustering as follows:

$$Degree\ of\ clustering = 1 - \frac{Median\ breaks'\ distance}{Maximum\ mean\ breaks'\ distance}. \tag{4}$$

This quantity is calculated for each realization of the experiment (depending on the number of breaks in each repetition) and then presented as mean of those individual values. The fraction itself can have the value from the interval (0,1], higher for breaks more distant from each other. After taking the difference from one the value of the measure changes in the opposite direction than the distance between the breaks, which better follows the notion of clustering. When the median distance decreases more than the potential maximum mean distance in a sample then the breaks are closer together, i.e. the degree of clustering is higher. The value of the statistic appears to be around one third on data with equal distribution of breaks. That fact stems from a relationship between mean and median in general and is not of interest for our research since we use the degree of clustering only as a relative measure.

There are several reasons why we are not using the probability itself as the explanatory variable for the estimated long memory parameter (even though it is only one number in case of the first hypothesis) instead of the artificially defined degree of clustering. First one follows the same logic as was described above for the definition of the maximum mean break distance. We want to compare the realized distance with the realized potential maximum, while p_1 is a theoretical value (even though we repeat the experiment sufficiently for the realized values to converge to their expected value). The Degree of clustering is in a sense an instrument for the parameter p_1 even though we know its true value. This fact connects to the second reason for not using p_1 directly – we believe that our simulation is thanks to this method closer to the real world situation when the structural breaks are not statistically given and the probability of occurrence is only an estimated value. It is possible to calculate the degree of clustering on real world data (after identifying the breaks by some method) and thus also to match the situation with its simulated version. Third reason is quite practical – the notion of degree of clustering as defined above is much more intuitive than the original concept of additional probability. This is important for the formulation of the hypotheses and comparison of the results with our expectations. Last but not least in more complex simulations (with substantially more parameters) further on in this thesis it is the only possible way of work so we also avoid inconsistency in employed explanatory variables (the Degree of clustering is comparable over different specifications of data generating processes).

2.4 Other Parameters of the DGP

There are several parameters which are not of interest to our research but still have to be set. In the beginning of the thesis we mentioned that for simplicity the value of shifts (meaning both their magnitude and sign) comes from a distribution with zero mean. In case of the first hypothesis' testing the only step we take to get closer to the real world situation is that these values of shifts come from a normal distribution. Another alternative which we have considered was for example the uniform distribution. According to our idea of "structural breaks" they are formed by a combination of many external factors so there is no reason their value should be distributed over a strict interval.

The variance of the distribution can be set arbitrarily; however it must correspond to the variance of the core process (for example the above mentioned ARMA). If it was not so, the process of shifts to mean would render the core process insignificant and no robustness checks could be done. We want to estimate the long memory parameter of some version of ARMA with shifts to mean, not a long memory parameter of a modified random walk (Equation 1). Suppose we have an ARMA(1,1) process with AR parameter α_1, MA parameter β_1 and innovations $\mu_t \sim \Phi(0, \sigma_\mu^2)$. The asymptotic variance of the process is

$$\sigma_{ARMA}^2 = Var\big(ARMA(1,1)\big) = \frac{(1 - 2 * \alpha_1 * \beta_1 + {\beta_1}^2)\sigma_\mu^2}{1 - {\alpha_1}^2}. \tag{5}$$

Based on Equation 1 the mean value of the cumulated shifts is zero (innovations come from a distribution with zero mean). Denoting the total probability of break at a given point in time by p and the number of time periods by n, the variance of the cumulated shifts is

$$\sigma_{Shifts}^2 = n * p^2 * \sigma_\varepsilon^2.$$

Supposing we want to keep the variance of shifts at the same value as the variance of the core process

$$\sigma_{Shifts}^2 = \sigma_{ARMA}^2.$$

Then we have to set the variance of distribution from which the shifts come to

$$\sigma_\varepsilon^2 = \frac{\sigma_{ARMA}^2}{n * p^2}. \tag{6}$$

Since the process on which we estimate the long memory parameter is the sum of the core process and of the cumulated shifts to mean its variance is $\sigma_{Shifts}^2 + \sigma_{ARMA}^2$.

3 The Data Generating Processes

3.1 First Hypothesis

Hypothesis 1: The estimated long memory parameter depends on the degree of structural breaks clustering.

The matter of interest for testing of the first hypothesis is the distribution of structural breaks within the sample and how changing parameters of the distribution affects the estimated long memory parameter of the resulting data (which is a combination of the initial process and the cumulated shifts to mean). As a starting point there is a given probability of the break happening, equal for all data points. First additional parameter affecting the distribution of breaks is the increase in probability of break occurrence at points in time immediately following another break, so the probability of a break happening in each point in time is the following

$$p_T = p_0 + p_1 * p_T. \tag{7}$$

In order to be able to assess the effect of changing this parameter (p_1) alone, the expected number of breaks in the sample must be the same for all its values. If it was not the case there could be a different estimated long memory parameter solely due to higher number of realized breaks. Therefore the total probability of break p_T is a parameter of the model and the independent probability p_0 needs to be calculated – the formula

$$p_0 = p_T * (1 - p_1). \tag{8}$$

determines its value for each step (we are linearly increasing additional probability p_1 from zero up to its maximal value, see part 2.2 for more detailed explanation of the concept of steps). We are not increasing the probability of a break happening – we are rather unbalancing the situation, shifting the probability towards points in time following another break (whether induced solely by the p_0 component or already by the additional p_1).

3.2 Second Hypothesis

Hypothesis 2: The estimated long memory parameter depends on the lag of structural breaks clustering.

Major part of the approach remains the same – starting point is the core process (AR, MA, ARMA) to which we add the cumulated shift to mean. What changes is how the break points at which we cumulate the shifts are determined. During testing of the first hypothesis we simply cared about whether a break has occurred in the previous period. In other words the maximum lag of break dependence was equal to one. In the part our research dedicated to testing of the second hypothesis we focus on what happens to the estimated long memory parameter if we increase the maximum lag of structural breaks' dependence.

The idea behind the second hypothesis is that – except for further increasing the degree of clustering of the simulated data – higher lag introduces truly the "long term dependence" from the definition of long memory. Simulating the data in such particular way aiming at the definition of the long memory may seem artificial; however we are at the same time still trying to be in line with the real world situation. It is understandable that if there is a certain level of structural breaks' clustering present in the real world data then the dependence is not only of lag one. Since in reality the mean of a time series can fluctuate over several periods – not only two – it would imply an extremely high parameter p_1 if the situation was to be parallel to the state described in the section 3.1 (with the maximum lag equal to one).

Based on this discussion we believe we are free to experiment with the values of higher term dependencies – parameters p_2, p_3, ..., p_k from the equation

$$p_T = p_0 + p_1 * p_T + p_2 * p_T + p_3 * p_T + \cdots + p_k * p_T, \qquad k < n, \tag{9}$$

which is only the Equation 7 extended for the lag up to length k. Given this specification we need to consider the situation in two respects: First, how much backward-looking should the distribution of breaks be (i.e. determine the value of k). Second, how fast should the parameters p_2, p_3, ..., p_k decay. Changes in these two characteristics will be examined as for their impact on the long memory parameter estimated on the simulated data.

A rule parallel to Equation 8 must hold in order to keep the expected number of breaks equal no matter which values we assign to parameters $p_2, p_3, ..., p_k$. Setting up the rule

$$\sum_{i=1}^{k} p_i \leq 1, \tag{10}$$

we get the formula

$$p_0 = p_T * (1 - p_1 - p_2 - p_3 - \cdots - p_k). \tag{11}$$

Thus the value of p_0, the independent probability of break, is the residual value of the total (constant) probability of and the unbalancing parameters $p_2, p_3, ..., p_k$. We can understand these unbalancing parameters as a certain kind of measure for the persistence of transition period. If we project some more complex patterns into their values we can for example model some inner cyclicality within the periods of crises.

The simplest variant of pattern in the vector of additional probabilities is an even distribution over all lags. It means that we give an equal importance to all past periods in consideration – if a break occurred in any of them it will have the same effect on the probability of occurrence in the current time period:

$$p_i = \frac{Added\,probability}{k}, i = 1 \ldots k. \tag{12}$$

In most applications of any measures evaluating the past, the earlier periods are "weighted out" in some sense. The following equation serves to determine the probability added in case a break has occurred i-th lag before, dividing the total added probability into a sequence linearly decreasing over k lags:

$$p_i = \frac{Added\,probability}{k} * \frac{k-i}{k}, i = 1 \ldots k. \tag{13}$$

The pattern with equal additional probabilities is a good starting point for our experiment as we can examine how the outcome will develop as we change it to something more realistic. An interesting discussion is what effect could have adding another lag into this pattern compared to a pattern where the past is weighted out. On one hand, we could expect that it will be larger for the equally distributed pattern simply because the relevant probability parameter being added in another lag is larger. On the other, the probability parameters in all of the previous lags are in both cases equally decreased by a fraction of total added probability equal to inverse of the new number of lags:

$$Added\ probability = \sum_{i=1}^{k} p_i,$$

$$k \to k+1: \forall i \in [1, k-1]: \Delta p_i = -\frac{Added\ probability}{k}.$$

This second view suggests that the effect of adding another lag will be the same for both discussed patterns.

There is yet another alternative which was already mentioned before – a more complex pattern in which higher importance is given to higher lags than to lower. It could be explained as introducing some sort of cyclicality into a financial turmoil – imagine a situation when there is a constant delay in the accessibility of information about market conditions but the trades are realized in real time. Market prices can reflect a structural break only when the market agents know about it, thus with this delay. We will not simulate this pattern with an additional probability higher strictly for the one past period but gradually decreasing up until the first lag:

$$p_i = \frac{Added\ probability}{k} * \frac{2i}{k}, i = 1 \dots k. \tag{14}$$

With this rule efficiently achieve an inverse situation to when the past is linearly weighted out (Equation 13) and if there are any differences in the outcome it will be much easier to quantify and interpret them, while the idea of delayed information will still remain expressed by the pattern.

The magnitude of breaks at the simulated break points still remains independent and comes from a normal distribution with zero mean and variance set by Equation 6. The basic core process used for the testing of the second hypothesis is an ARMA(α_1= -0.1, α_2= 0.05, β_1 = 0.1, β_2= 0.05) process. The results are checked for robustness by applying different core processes.

3.3 Third Hypothesis

Hypothesis 3: Clustering of same-sign shifts to mean has different effects on the estimated long memory parameter than sign-varying fluctuations.

3.3.1 Manipulating the Direction of Shifts

The third hypothesis can be seen as an extension to the second hypothesis. We add more lags into the data generating process of the first hypothesis like we did for the second one. At the same time we take into consideration the sign of the shifts occurring at past break points up until the highest lag. To express the overall change over the lagged periods we calculate the mean shift size. It is important to stress that we do not calculate the mean of given number of previous shifts but the mean size of shifts during the given number of preceding periods. The reason for this is that the distance in time between two shifts can be relatively high and we want to control the "memory" of the additional probability by the highest lag used. Our idea also is that in such setting the estimated long memory parameter will be more sensitive to the number of lags: the stronger is the clustering of breaks, the more of them will fit into the period of given number of lags. We expect that increasing the number of lags – providing stronger memory – will cause the direction of shifts to be more persistent in the case of the same-sign shifts and less persistent in the case of varying sign. We still do not examine this persistency directly but only in its effect on the estimated long memory parameter.

For testing of the third hypothesis, all simulations are run for the scenario with same-sign shifts and the resulting estimates are compared to those from testing of the second hypothesis. Then the same is done for the case of sign varying shifts and the differences in development over the steps are interpreted. We always compare the results from the simulations with the same number of lags.

The data generating process of shifts to mean for the scenario with same-sign shifts takes the following form:

$$
\begin{gathered}
Shift_t = Shift_{t-1} + m_t * Break_t \\
Break_t = \begin{cases} 1, & with\, p_t \\ 0, & otherwise \end{cases} \\
p_t = p_0 + p_1 * Break_{t-1} + p_2 * Break_{t-2} + \cdots + p_k * Break_{t-k} \\
m_t = sgn\left(\frac{1}{k}\sum_{i=1}^{k} m_{t-i}\right) * |\varepsilon_t|, \qquad \varepsilon_t \sim \Phi(0, \sigma_\varepsilon^2)
\end{gathered}
\tag{15}
$$

where the term m_t stands for the magnitude of shift at time t. It is clear that the size of the shift to mean itself still comes from a normal distribution but the direction of the change is modified. In case there was no break over the course of last k time periods the sign of the shift comes from the normal distribution as well. This shows again that in case of evenly distributed breaks the results should not be significantly different from the results of testing of the second hypothesis. We expect to see this similarity for the lowest steps when the additional probability is very low.

In contrast to all previous data generating processes the one described in Equation 15 has a non-zero mean. In fact with the number of time periods approaching infinity it would diverge with the speed of divergence depending on the parameter k and additional probabilities p. Once the clustering is substantial and the memory of the distribution of breaks strong enough the magnitude of the next shift will never "escape" the direction of previous shifts. Cumulating the shifts with the same sign means the divergence is linear. The concept of long memory can be expressed as (and is measured as) the presence of some low frequency waves within the data. So far with the direction of shift being completely random the structural breaks were seemingly causing such waves. Because of that we do not expect the estimated long memory parameter to react significantly to changes in the distribution of breaks in this case, only for lower number of lags.

The simplest way to achieve the sign varying shifts would be to reverse the sign of the previous shift. But as was said in the beginning of this subsection we want to be able to control the memory of the breaks' distribution so we will never look at last k breaks, rather at the last k time periods. The simplest approach has one more weakness – imagine a situation when by chance all even breaks in some sequence are larger and positive and the odd ones are smaller and negative. In the large sample these will on average equal but we want to magnify this property – speed it up in a form of mean-reversal. Because of that we will use the specification from Equation 15 which is already backward looking at the mean of shifts during past periods and make only a small adjustment:

$$m_t = -sgn\left(\frac{1}{k}\sum_{i=1}^{k} m_{t-i}\right) * |\varepsilon_t|, \qquad \varepsilon_t \sim \Phi(0, \sigma_\varepsilon^2) \tag{16}$$

This process produces breaks with random magnitude but with the sign opposite to the sign of the mean shift size over the last k periods.

Unlike our expectation regarding the consequences of the process described in Equation 15, it is quite more complicated to imagine the effect of the variant from Equation 16. We see as one possibility that the effect can be exactly the opposite of the process with the same-sign shifts: After a period with a positive mean shift size only negative shifts will be generated until one of the following happens – either the positive shifts are overweighed by the magnitude of the negative ones, or the positive ones are "forgotten" – they exit the window of considered periods. Then the positive shifts can prevail again and the whole situation repeats. That would imply high estimated long memory parameter – the actual structural break does not happen when the $Break_t$ in Equation 15 is equal to one but when the sign of the mean shift over the last k periods changes, which would introduce very slow waves into the data pattern.

The second possibility is that the switching of the sign reduces the estimated long memory parameter. That could be the case if the sign changes come too often. A regularly oscillating break does no longer represent a structural change and it becomes a part of the core process, indistinguishable from the ARMA waves.

Because we add more lags into the data generating process as we do for the testing of the second hypothesis we need to decide which pattern will the vector of probabilities at particular lags follow. Disregarding results from simulation with the individual patterns we incline towards using the probabilities linearly decreasing into the past. It is so because we believe this pattern is simple yet close to most of the real world situations. We follow an arbitrary rule for the variance of shifts' size σ_ε^2, setting it equal to one fifth of the variance of innovations to the core process. Why we set up this rule is explained further on in the section 4.2 dedicated to the results of testing of the second hypothesis. The basic core process remains ARMA(α_1= -0.1, α_2= 0.05, β_1 = 0.1, β_2= 0.05) although we perform robustness checks with different processes described in the Chapter 4.

3.3.2 Manipulating the Size of Shifts

In addition to the data generating processes manipulating the direction of shifts described above we test an extension to the third hypothesis. The same intuition as for tampering with the sign of breaks' magnitude can be applied to how we determine their size in general. The idea comes once again from the properties of the real world data, especially from how they behave during crises. Namely we try to imitate the empirically confirmed phenomenon of volatility clustering. Since we already examine the clustering of breaks itself, introducing a dependency between their sizes in consecutive time periods is at hand. The volatility clustering is explained by Satchell and Knight (2011) as following:

Volatility clustering. The [...] stylized fact is the clustering of periods of volatility, i.e. large movements followed by further large movements. Correlograms and corresponding Box-Ljung statistics show significant correlations which exist at extended lag lengths. (Satchell and Knight, 2011)

Let us think about what this property of data means for the long memory. Let us imagine a situation when a large structural break occurs at some point in time. Due to the clustering of breaks introduced by us to the process it is more probable that more breaks will happen in the subsequent periods. These breaks are also relatively large according to the rule of volatility clustering. However, the rule defined above says nothing about the direction of the shifts. It describes a situation when market agents are more sensitive to new information – under normal conditions in a calm period the fluctuations would on average offset each other and leave only the trend. It is the same during a turmoil with increased volatility – only both of the offsetting powers are stronger. Therefore two states with different implications for the estimated long memory can follow. First is the one in which the clustered breaks are balanced as for their direction. The cumulative shift to mean strongly fluctuates and then returns close to its previous value. A high number of large changes during a short period of time cannot be mistakenly interpreted as a sign of long memory by any estimator. Estimation of such process would probably aim at some form of GARCH model. The second possible state following a large structural change is a situation when the signs of the shifts are not well balanced. Since the clustered shifts are large the shift in cumulated mean is also large. This set of events can repeat until there is a cluster of small shifts following a cluster of large shifts. This final combination forms on bigger scale one wave with very low frequency which can be interpreted as a sign of the long memory present in the data. Important for the results is what state will occur more frequently and how is it influenced by the number of lags or any other parameters of our models.

The simulations with volatility clustering represent a final stage of our research in this thesis. They combine or build on the simulations done in other parts of this chapter. As in part 3.1 we manipulate the probability of a break, as in part 3.2 we introduce a dependency on multiple lags with various patterns in the vector of additional probabilities, and as in the part 3.3.1 we set a rule for the magnitude of shifts as well. However in the previous part the rule set for the sign of the shifts' magnitude was very strict – it did not depend on any other external probabilities than those which were present already in the previous stages of our experiment (i.e. solely on values from the normal distribution which was the source for the shifts' size so far). It does not matter whether we consider the case of the same sign shifts or the case with alternating ones, the data generating processes defined in Equations 15 and 16 always yield a clear instruction for what the sign of the new shift is based on information from previous periods. In the case of volatility clustering we do not want it to be so. What we need is a process in which the magnitudes are dependent with certain probability, meaning that depending on whether a shift is large or small it is more probable for the next one to be also large or small, respectively. We stress the term “more probable” because that is exactly what is described in the definition of the volatility clustering, which put to simple words says that large breaks tend to follow large breaks. If it was not so and the magnitude would follow rules similarly strict as the ones in the Equations 18 and 19 the size of the shifts would end up trapped on certain level. One way to achieve this dependence in probability would be to introduce another vector of probabilities to influence the magnitude of future breaks and scale the vector based on the size of past breaks. However introducing another set of probabilities into our model is not what we favour – it would require setting even more rules as to how sensitive the scaling based on previous breaks’ magnitude should be and how large should the probabilities be in general.

The idea how to deal with this problem comes from a complication with which we had to deal when simulating the data for testing of the second hypothesis, the part of the experiment dedicated to the pattern of equally distributed probabilities. We have discovered that the rule for the variance of the probability distribution from which the size of the breaks comes, set in Equation 6, causes the estimated long memory parameter to be too close to its upper limit in case of the second hypothesis’ testing. Simply put the shifts generated according to this rule are too large in relation to the core process. Despite how the rule had to be adjusted for the results of the testing to be valid we can use it again to manipulate the magnitude of shifts in particular periods, not in the general setting of the data generating process, to achieve volatility clustering in the resulting data. Introducing the time dimension into the rule for volatility of shifts gives us the starting point for the data generating process:

$$m_t = \varepsilon_t, \qquad \varepsilon_t \sim \Phi(0, \sigma^2_{\varepsilon,t}) \tag{17}$$

Now we need to determine the rule setting the parameter $\sigma^2_{\varepsilon,t}$. The immediate approach would be to simply use some variant of an ARCH or GARCH model as it is done in other researches and suggested by many articles (Poon and Granger, 2003; Hansen and Lunde, 2005; Shepard, 1996). However the problem with these models is that they are designed for modelling of the change of an underlying variable in all time periods. Compared to such situation our structural breaks are relatively rare – according to our models they occur on average only in 5% of time periods (based on the parameter of total probability of break's occurrence), and it makes sense to interconnect their magnitudes only thanks to the fact that they are clustered together in time. Let us consider the following basic ARCH model specification:

$$\varepsilon_t = \sigma_t z_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{k} \alpha_i \, \varepsilon_{t-i}^2$$

As was previously said, the problem is that this model takes into account all time periods. Therefore we make a small adjustment so that we look at past periods only if a break happened in them.

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{k} \alpha_i \, \varepsilon_{t-i}^2 * Break_{i-t} \tag{18}$$

The model becomes an ARCH with time-varying coefficient where the coefficient at time t is equal to $\alpha_i * Break_{i-t}$. The parameter k corresponds to our number of lags (a parameter we want to influence so that we are able to compare the results to those from the testing of other hypotheses), the coefficients capture the dependency as we need it. Their values are set so that we stress the magnitude of breaks which happened at the most recent lags at the cost of the size of the shifts at the earlier lags. As was stated before, we need the rule to take into account at the same time both of the following – which lag the determining past break happened in, and how large it was in relation to its distribution. From these two needs the coefficients in the Equation 18 fulfil the former one. The latter one is captured by the model specification itself. The complete specification of the data generating process of shifts therefore is:

$$Shift_t = Shift_{t-1} + m_t * Break_t$$

$$Break_t = \begin{cases} 1, & with\, p_t \\ 0, & otherwise \end{cases}$$

$$p_t = p_0 + p_1 * Break_{t-1} + p_2 * Break_{t-2} + \cdots + p_k * Break_{t-k}$$

$$m_t = \sigma_t z_t, \qquad z_t \sim \Phi(0, 0.2 * \sigma_{ARMA}^2) \tag{19}$$

$$\sigma_t^2 = 1 + \sum_{i=1}^{k} \alpha_i \, m_{t-i}^2 * Break_{t-i}$$

$$\alpha_i = -0.5 + \frac{i-1}{k}$$

We can see that in case there were no breaks in the previous k periods the magnitude of the break in the current period comes from the exactly same distribution as it was for the testing of the second hypothesis. The approach to selecting the core process is the same as in all previous cases – the experiment is initially run with an ARMA(α_1= -0.1, α_2= 0.05, β_1 = 0.1, β_2= 0.05) core process and subsequently checked for robustness with different specifications.

4 Results

4.1 Testing of the First Hypothesis

Testing is performed on a sample of length n = 1000. The total probability of a break at each point of time $t \in [1, n]$ is set to p_T = 0.05. The ceiling for additional probability of break following another break is set to p_1^{max} = 0.1. The independent probability of break p_0 follows the rule set in Equation 8 and lies on interval $[0.027, 0.030)$ for values of $p_1 \in (0, p_1^{max}]$.

In Figure 2 we present an example of breaks distribution for values of the parameter p_1 equal to 0.005 (a), 0.035 (b), 0.070 (c) and 0.100 (d). It is observable by close look that the breaks are getting clustered together – increasing number of thicker lines and longer white spaces. After one thousand repetitions the mean number of breaks in a sample is 49.85 which is very close to the expected 50 breaks per realization.

Figure 3 shows the trend of increasing degree of clustering in terms of measure from Equation 4 averaged over one thousand realizations. As was mentioned before with balanced probability of break in each period (p_1 = 0) the degree of clustering corresponds approximately to one third. Gradually increasing the parameter p_1 up to 0.1 rises our measure on average by 0.15. This confirms that as for its value our measure is not very sensitive in absolute terms. It still performs well in relative terms; the upward slope is clearly visible and also confirmed by a simple linear regression

$$Degree\ of\ clustering_i = \gamma_0 + \gamma_1 * p_{1,i} + u_i \tag{20}$$

which yields an estimate of $\hat{\gamma}_1$ = 0.33 significant on 1% level and adjusted R^2 = 0.88.

Having simulated the process of breaks we focus on the values of shifts which happen at these points. Following the discussion for derivation of Equation 6 we first have to decide about what process to use as the core. First one is an ARMA(1,1) process with drift term μ = 2, AR coefficient α_1 = 0.10 and MA coefficient β_1 = -0.05 and innovations $u_t \sim \phi(0, 0.5)$. The asymptotic variance of this process is equal to 0.256. By Equation 6 we calculate that the desired variance of the distribution of breaks' value is equal to 0.284. Thus the Equations 1 and 8 take the form

$$Shift_t = Shift_{t-1} + \varepsilon_t * Break_t, \qquad \varepsilon_t \sim \mathrm{N}(0,0.284)$$

$$Break_t = \begin{cases} 1, & with\, p_t = \, p_0 + p_1 * Break_{t-1} \\ 0, & otherwise \end{cases} \tag{21}$$

$$p_0 = 0.03 * (1 - p_1), \qquad p_1 \in (0,0.1]$$

Figure 4 shows a vector of values of shifts with marked points of break from the example depicted in Figure 2. Figure 5 shows the resulting cumulated shifts to mean for parameter p_1 equal to 0.005 (a), 0.0350 (b), 0.0700 (c) and 0.1000 (d). It might seem from this example that the cumulated shifts tend to deviate from zero mean, however it is not so – evaluating one thousand realizations shows the mean of cumulated shifts to be almost zero (-0.003) and the mean value of shifts at break (ε_t for such t that $Break_t = 1$) is even closer to zero.

In Figure 6 we present the simulated ARMA(1,1) process with overlay of cumulated shifts to mean from Figure 5. Figure 7 joins these two parts together, showing the complete version of the simulated data. The waves characteristic for data with long memory property are clearly present in the sample; however it is not possible to tell without proper estimation how much they vary with changing parameter p_1.

Having simulated the process with desired properties we perform the estimation of long memory parameter d using the exact local Whittle estimator with parameter m = 0.4 corresponding to truncation at frequency of length of 200 observations. The estimate for varying values of p_1 is presented in the Figure 8. It is quite obvious already from the graph that the estimate has no significant relation to the increasing degree of clustering. To confirm it we run the simple linear regression

$$(Estimated\ parameter\ d)_i = \delta_0 + \delta_1 * (Degree\ of\ clustering)_i + u_i \tag{22}$$

The estimated value of coefficient $\delta_1$is -0.03, it is however insignificant even at very high levels and this regression has R^2 = 0.04 which is very low considering all other factors except the parameter p_1 were fixed.

To ensure robustness of the result we run the experiment with the same specification as above except for changing the core process to AR(2) with coefficients α_1 = 0.2 and α_2 = 0.1. The unconditional variance of this process is 0.266, implied variance of distribution of shifts 0.295. The resulting process with cumulative shift to mean is shown in the Figure 9 – there is no apparent difference compared to data based on the previous core process. The estimated coefficient δ_1 from Equaton 11 is -0.05, again insignificant at very high levels.

With such results we did not expect any outcome in favour of the first hypothesis. However for an MA(2) core process with coefficients $\beta_1 = 0.1$ and $\beta_2 = 0.05$ ($\sigma_\varepsilon^2 = 0.56$) the resulting estimate of coefficient δ_1 is 0.071, suddenly significant at 5% level. See Table 1 for summary of the simple regression and the Figure 10 for the graphical representation. From the plot it is visible that the randomness still has a very strong effect – even the mean estimated parameter over one thousand realizations is relatively erratic. That is in line with the low $R^2 = 27.8\%$. On the other hand the upward sloping trend is also visible and the positive estimate of coefficient δ_1 makes sense since the degree of clustering is increasing with parameter p_1. To validate the results we run the experiment once more with core process MA(3) with coefficients $\beta_1 = 0.4$, $\beta_2 = -0.1$ and $\beta_3 = 0.05$ ($\sigma_\varepsilon^2 = 0.59$). The estimate of the coefficient δ_1 is 0.043, again significant at 5% level.

Based on the results described in this part we conclude that the evidence does not support the first hypothesis. Significant results were yielded only for a specific core process and they are not robust to any changes of the data generating process. At the same time we are neither able to reject the first hypothesis. Therefore we cannot say whether the estimated long memory parameter depends on the clustering of breaks in this simplest setting.

4.2 Testing of the Second Hypothesis

Testing is done on sample length of 1000 time periods in 200 repetitions, with total probability of break occurrence in each period equal to 5% and maximum added probability equal to 60% (over the course of all lags). Three subsections follow, one for each pattern of additional probabilities – see the description of data generating processes for the second hypothesis in the Methodology section.

4.2.1 Equally Distributed Additional Probabilities

We start with five lags of the probability dependence and Equation 12 implies

$$p_i = 0.12, i = 1 \dots 5$$

The implied degree of clustering and mean estimated long memory parameter are presented in the Figure 11 and the summary of the simple linear regression from the Equation 22 can be found in the Table 2.

The results are in favour of our second hypothesis with the explanatory variable significant at 1% level and R^2 higher than 50%; however we see one limitation of their validity. The problem is that the estimated long memory parameter is very close to its upper limit (0.5) for the whole range of probabilities. It could mean that the long memory present in the data is actually "stressed" against the maximum value, resulting in underestimation of the examined effect of clustering or in general resulting in behaviour of the estimate which is hard to explain only by means of different breaks occurrence. Our idea why this is happening is that the magnitude of shifts is simply too high. The magnitude is determined by the variance of the shift-generating distribution, which was in case of testing of the first hypothesis selected based on the rule described in Equation 6. Now it seems that trying to keep the variance of breaks at the same level as the variance of the core process can be misleading. To deal with this problem, we will connect the variance of breaks directly to the variance of innovations in the core process.

We set an arbitrary ratio to 0.2, i.e. the variance of shifts' magnitude is equal to one fifth of the variance of innovations to the core process. The degree of clustering remains unchanged – we have changed only the approach to determining the magnitude of shifts (decreased their variance). The mean estimated long memory parameter now ranges from 0.25 to 0.3 and with only two outliers from the raw long memory parameter (before taking mean over the repetitions) touching the 0.5 limit, which we consider acceptable. We further examine the effect of changing this ratio in connection to the trade-off between breaks' size and their number within the sample (i.e. the total probability of occurrence). The results from such testing are presented in the section 4.4.

The estimated coefficient δ_1 from Equation 22 is equal to 0.073. This coefficient indicates how much the long memory parameter would change if we switched from no clustering to perfect clustering. The degree of clustering increases approximately by 42% over the 20 steps of probabilities' scaling. If we add one lag, implying $p_i = 0.1$ for all lags we get the results presented in Table 3. The estimate of the coefficient δ_1 is now even lower which goes against our intuition and expectations. However a sample of 1000 time period is relatively large to examine an effect of adding only one lag. Because of that we gradually increase the number of lags up to 10 and also test the situation for 15 and 20 lags. Again it is important to stress that we are not increasing the total added probability – it remains constant on the level of 60%. In light of this fact the overall increase of the estimate of coefficient δ_1 which is presented in Table 3 is really caused solely by "prolonging" the memory of breaks' distribution process. If we compare the estimate from situation with 5 lags with the one from simulation with 20 lags, the increase is substantial. Even though we are quadrupling the number of lags, in terms of the sample it means adding only 1.5% of the sample length, and it leads to increase in the estimated coefficient by 0.1. These results imply that adding more lags to the dependence of breaks' occurrence in an equally distributed pattern has a significant effect on the estimated long memory property seemingly present in the data.

4.2.2 Reducing the Importance of the Past

The distribution of additional probabilities which puts more stress on the most recent periods seems less arbitrary than the one with equal distribution. First instance contains again five lags of the probability dependence so from Equation 13 we see that the additional probability of break occurrence immediately after another break (i=1) is 20% and it is linearly decreasing over the next four lags. The degree of clustering implied by such probabilities and the estimated long memory parameter are presented in Figure 12. The summary of simple linear regression as described in Equation 22 can be seen in Table 4.

The estimated coefficient δ_1 from Equation 22 is equal to 0.049 and it is significant at 1% level. We need to remember that our degree of clustering is expressed on a scale from 0 to 1. As was stated before, the degree of clustering as we define it is a measure for how much we manipulate the breaks' distribution. In order to interpret this result we need to keep in mind how we were adjusting it in this specific case. We can say that increase in the degree of clustering by 38 percentage points induced by dependence of breaks' occurrence over five lags increases the long memory parameter by approximately 0.049, which is not a very strong effect.

If we repeat the same experiment but increase the maximum lag in dependence of breaks' occurrence to 6 (Table 5), the newly estimated coefficient $\delta_1$is equal to 0.074 still with the same significance and even higher R^2 approximately equal to 78%. That is a major increase which motivates us to further increasing the lags. Table 6 shows an overview of the estimations of coefficient δ_1 with increasing number of lags of dependence. At ten lags we see that doubling the period of dependence in its length has also doubled the estimated effect of volatility clustering on the estimated long memory parameter. At 15 lags is the increase in the estimated coefficient still as high. When we compare the situation at 20 lags with the simulation with 5 lags like we did in case of equally distributed probabilities we see that the increase in the estimated coefficient δ_1 is now even higher. That is interesting since the importance of higher lags has decreased in relative terms, and such change means getting closer to the specification of the first hypothesis' testing. However we do not question the result as this pattern of probabilities distribution is, according to our view, closer to any real world example than the case of equal distribution.

4.2.3 Increased Importance of Earlier Periods

The results described in the sections 4.2.1 and 4.2.2 are in line with our expectations. The situation when higher lags are more determining for the occurrence of break than the lower lags (recent periods) is somehow opposite to the one with less important past. Starting again with five lags the probability added based on the period immediately preceding another break is 0.04 (from Equation 14) and it linearly increases up to 0.2 on the fifth lag. The degree of clustering created by these probabilities and the resulting estimate of the long memory parameter d are presented in Figure 13 and the summary of the regression according to Equation 22 is provided in Table 7. The estimate of the coefficient δ_1 is now 0.059, higher than in the previous subsection but still lower than for the case of equally distributed probabilities. Five lags are a very short part of the sample; we focus on the development of the estimate as we add more lags. This overview can be seen in Table 8. Overall, the estimate of the coefficient is increasing with adding more lags as it was in previous two cases. However the increase is relatively slower – at 20 lags the estimate is lower by 0.04 than in the situation with equally distributed additional probabilities and even by 0.11 than in the case of decreasing probabilities. Also in the part with less than 10 lags an additional lag does not always mean a higher estimated coefficient.

Based on the above described facts we conclude that in order to capture a delay in release of information or some cyclicality during crises, the pattern of additional probabilities would have to be a bit more complex, possibly with multiple peaks and with a gradual decrease over the highest lags – it would have to be a combination of the patterns tested in the sections 4.2.1 and 4.2.2. We will still use the results of this subsection as a comparative to sections 2.2.1 and 2.2.2; however we will not try to make any general conclusions or assessments of the real-world applications based on the numbers.

4.2.4 Comparison of the Three Patterns

The overall development of the estimate of the coefficient δ_1 with increasing number of lags is similar for all three examined patterns – the estimate increases with higher number of lags. Also this increase is not very smooth in any of the cases and some drops have appeared, but only in terms of adding one lag – not across a higher number of additional lags. On the level of 20 lags, which was the highest examined number, the estimated coefficient is substantially higher for the second pattern (Equation 13) and lowest for the third one (Equation 14).

From the results described in the three subsections above and summarized in Tables 3, 6 and 8 we conclude that we do not reject the second hypothesis as we have found a significant dependence between the number of lags of structural breaks' clustering and the estimated long memory parameter. Real world representation of this conclusion would be a persistent crisis or in general a period of lasting uncertainty in the market. In such situation a market agent is changing the system of his behaviour in reaction to the past behaviour of other market agents. Our result implies that the long memory parameter estimated on the data from such market would be higher even if a fractionally integrated data generating process was not present. Forecasting based on historical data generated under such conditions must be resistant to overestimation of the long memory parameter.

The core process employed for the testing of the second hypothesis is ARMA(α_1= -0.1, α_2= 0.05, β_1 = 0.1, β_2= 0.05). To check for robustness against the change of the core process we re-run all of the previous simulations from the parts 4.2.1, 4.2.2 and 4.2.3 with the following two core processes: AR(α_1 = 0.2, α_2 =-0.05) and MA(β_1 = 0.3, β_2 = 0.1, β_3 = -0.05). The results are presented in Table 9. The estimated coefficients vary only slightly from the previous results and the development with increasing of the number of lags is the same. Based on this we conclude that the core process which we chose was not significantly influential to the estimated parameters and coefficients and the results presented in this section are valid without loss of generality.

4.2.5 Variance of the Estimates

A possible argument against the validity of our results could be that the variance of the long memory parameter estimates is too high. In the section 'Estimating the long memory parameter' we discuss what truncation parameter should be used in the Exact Local Whittle estimator to balance the trade-off between bias and variance. In all of the simulations so far we used the truncation at one fifth of the sample length. Since the variance of the estimates seems to be too high from the box plots (Figures 12, 13) we re-run the case of second hypothesis with its second (decreasing) pattern with 10 lags with truncation at two fifths of the sample (400 periods out of the 1000 total sample length). See Figure 1 for evaluation of the variance-bias trade-off. The resulting estimates of the long memory parameter with higher truncation are shown in Figure 14. The decrease of variance can be seen even from the box plot; numerically is the mean variance over the 20 steps 0.0059 in the original case with lower truncation and 0.0034 in the case with higher truncation. On the other hand also a decrease in the estimate of parameter d is visible from the plot for all of the steps and the estimated coefficient δ_1 has dropped from 0.108 to 0.063. Since we can deal with the problem of high variance by repeating the simulation sufficient number of times we will rather try to avoid the underestimation of the long memory parameter and keep the truncation at the lower level.

Summarizing the results from the part 4.2 we say that there is substantial evidence of the effect of the number of lags on the estimated long memory parameter for us not to reject the second hypothesis. The effect is significant for all the different settings and varies only in the strength depending on the pattern in the additional probabilities' vector.

4.3 Testing of the Third Hypothesis

4.3.1 Manipulating the Direction of Shifts

The sample length, number of repetitions, total probability of a break in each period and the maximum added probability over the course of 20 steps remain the same as they were for the testing of the second hypothesis.

First we focus on the situation with same-sign shifts. The degree of clustering resulting from running the model specified in the Equation 15 with maximum number of lags equal to 5, together with the long memory parameter estimated on the generated data, is presented in Figure 15. The summary from the regression according to Equation 22 can be found in Table 10. Already from the first look it is obvious that the situation is completely different from all of the previous simulations and especially completely opposite to what were our expectations regarding this result. The estimate of the coefficient δ_1 is equal to 0.237, much higher than in all of the previous instances of our experiment. This means that introducing the cumulative shift to mean with higher probability of the same-sign shifts (this description is very simplified, for full definition see the part 3.3 dedicated to data generating process of third hypothesis) has much stronger effect on the estimated long memory property than a process with no regards to the sign or size of shifts (i.e. with zero mean, as for the first and second hypotheses). Also when we increase the number of lags the increase in the estimate of the coefficient is very steep. The overview of all of the estimates with respect to the maximum number of lags is presented in Table 11. At 20 lags, the maximum number for which we run the simulations, the estimate is even as high as 0.646. The interpretation of this coefficient is that in comparison with the situation free of any interference with the breaks distribution, in a situation with additional probability of break reaching 20 periods in the past the estimated long memory parameter is higher by 0.00646 for each 1% increase in the degree of clustering. Since the degree of clustering increased in our simulation approximately by 14% the resulting overestimation of the long memory parameter nears 0.1 – which is a considerable influence.

This finding means that the real world data from a situation when the same-sign shifts are more difficult to predict with precision in the presence of long memory. In order to create an accurate forecast the data generating process must be determined correctly, including the long memory parameter. As we see in our result this parameter can get easily and substantially overestimated in such setting.

In order to understand why the outcome of the experiment is so surprising compared to our expectation – we expected very low influence on the estimated long memory parameter – we take a closer look at the simulated size of shifts and final data after joining with the core process. Figure 16 shows the shifts during the whole sample together with the respective final process in three situations, all from simulation with 10 lags. First situation is the one with the Step equal to 1 (no tampering with the clustering, only with the sign), the last one for Step 20 (the additional probability reaching the maximum from all simulations, in the specific case 60% distributed over the 20 lags), and the middle one for Step 10 – the midpoint between the other two. We see that the divergence which we expected for reasons described in the part 3.3.1 is either not present at all or extremely slow. The reason seems to be that since the breaks are getting more clustered together with higher Step there is also more time between individual clusters. During this time the maximum lag is surpassed and the sign of all of the previous shifts is "forgotten". Outside of the sign-setting rule a reversal of the direction of shifts can happen and during the next cluster the shifts counteract the cumulative change created during the current cluster. The divergence therefore occurs only with low probability and if it is not the case, one large break is created per each cluster. In fact this joining of clustered breaks into one effect is what we anticipated already in the project of our research as a general outcome of some of the simulations yet we were not able to expect it to be outcome of this specific instance. It is still our opinion that with sufficient number of breaks occurring during the whole data sample the shifts would never get out of the same-sign trap and the effect on the estimated long memory parameter would be exactly the opposite. There is no need of testing it since a high number of breaks goes against the idea of structural change.

The data generating process for the case with sign varying shifts which is set in Equation 16 yields also an interesting outcome. The resulting degree of clustering, together with the estimate of the long memory parameter, is presented in Figure 17. Table 12 shows the summary from the related estimation of model according to the Equation 22. On contrary to the significant and very strong relationship discovered for the process with same-sign shifts the sign varying (or also mean-shift-size reversing) simulation finds no significant relationship between the increasing additional probabilities of break and the estimated long memory parameter. Furthermore the findings are consistent for all levels of maximum lag number, for comparison see the Table 13.

These results are not so surprising anymore since we already know what effects the same-sign shifts have on the simulated data and estimated long memory. Now when we use a process which in a way manipulates the data in the exact opposite way we understand why it has such consequences. On the scale of the whole data sample two shifts to mean with opposite sign occurring in close proximity offset each other. Together they have an effect on the estimated parameter d as would have a shift with size equal to the difference between their absolutes. Anytime such structural break happens that it moves the time series from its historical mean all subsequent shifts will have the opposite sign until the previous one is balanced out. The problem we examine is the interchangeability of structural breaks and the actual long memory present in the data based on empirical evidence. There is no danger of such confusion when the structural breaks occur with varying signs. Therefore it cannot be said in general that when structural breaks are present in a data sample the model must be corrected for the bias caused by the overestimation of the long memory parameter. First it is necessary to examine whether these breaks join together into some stronger effect or whether they balance each other out.

In Figure 18 you can see a comparison of two final simulated processes (i.e. the core process together with the cumulative shift to mean), both from the case with 5 maximum lags and from the last step of the simulation, thus with the highest expected degree of clustering. First process comes from the sign-varying scenario and its degree of clustering is 69%, the second is from the same-sign shifts' scenario with resulting degree of clustering at 74%. The estimated long memory parameter d is equal to 0.265 and 0.321, respectively. From the estimates of coefficient δ_1 presented earlier we see that the difference in the estimated long memory parameter is higher than which would correspond to the difference in the degree of clustering. That is exactly the effect of the manipulated signs. Moreover the same can be noticed already from the shape of the lines of these two processes – the same-sign process is clearly wavier while the sign-varying one shows mostly only the fluctuations of the core process.

The robustness check was done using the two processes AR(α_1= 0.2, α_1 = -0.05) and MA(β_1 = 0.3, β_2 = 0.1, β_3 = -0.05). The results indicate the same logic as was described for the original core ARMA process; numerically there is also no significant difference.

Analysis of the autocorrelation function

As the outcome of this part of the experiment is very important for our conclusions we want to support it with even more evidence. Based on the results presented in this section we say that clustering of the same-sign shifts increases the overestimate of the long memory parameter while the clustering of the sign-varying shifts decreases it and makes it independent on the maximum number of lags. To prove our point we compare the simulated final processes (a core process together with the cumulative shift to mean) with another simulation – an actual fractionally differentiated process, ARFIMA, with the AR and MA coefficients identical to the core process. We understandably set the parameter of fractional integration (or in other words the long memory parameter we are examining) in the simulated ARFIMA process equal to the mean of our estimates over the 200 repetitions. We compare the fractionally integrated process with those simulations with the "longest memory", i.e. the maximum number of lags equal to 20. We use the estimates from Step 20 since we want the manipulated probabilities to have the full effect. The mean estimate of the parameter d represented by the red line in Figures 15 and 17 and is equal to 0.398 and 0.181 for the same-sign shifts' and sign-varying shifts' case, respectively. According to its definition, the long memory is represented by a slow decay in the autocorrelation between two observations with an increasing time distance. Therefore the tool we use to support our findings is the autocorrelation function.

The comparison for the same-sign shifts can be seen in Figure 19. There is certain similarity in the shape of the series themselves – both suggest the existence of more complex process than a simple ARMA. Yet there is a difference in the length of the low frequency waves. They still seem to be shorter in the case of the real ARFIMA process than what is the effect of structural breaks. The slow decay is obvious in the autocorrelation functions of both of them but again a little different. The decline is more smooth and gradual for the fractionally integrated process. In case of the clustered process there is a substantial fall after the second lag (which is high due to the AR part of the process) and then it appears to be almost constant. That is exactly what indicates that the evidence of the long memory is false. According to its definition the long memory means a slow decay in autocorrelations while it is rather a constant property of the clustered process that the autocorrelations are too high.

In the next part of the figure we show the autocorrelation function of the two processes after fractional differentiation. It is very important to stress that we do not differentiate the processes to the degree suggested by the estimate of the parameter d (in case of the ARFIMA process we know the exact value). Instead we use a value lower by 25%. Based on the resulting autocorrelation functions we would say that the process with clustering is free of any signs of the long memory. The spikes of the autocorrelation function of the differentiated ARFIMA process are still too high. It implies that the parameter d was indeed overestimated in case of the process with clustering. For comparison we also show the autocorrelation functions of the core ARMA process and of the ARFIMA process differentiated by the correct fraction.

The Figure 20 documents the same procedure for the process with sign-varying shifts. Already from the plot of the process with structural breaks we can see the difference in comparison with the case with same-sign shifts. It is almost impossible to tell in which periods the breaks happen solely from the plot of the process (the jumps of the final process at the marked breaks are easily interchangeable with the jumps of the core ARMA process itself). The process is also even more similar to the fractionally integrated one, the notable difference in low frequency waves' length is no longer present. As for the subsequent steps of the process, the major difference is that differentiating by the lowered estimate of the parameter d is not sufficient and both autocorrelation functions still show the signs of the long memory property. It can be fixed by differencing both time series by the full value of the parameter d according to its estimate and simulated value for the ARFIMA process. The implications of such outcome are the same as those described in the section dedicated to the sign-varying shifts in general. The mean reversal represented by the switching of the sign causes the data to produce evidence of long memory property. Compared to the same-sign shifts the estimated parameter d is much lower and it does not depend on the maximum number of lags, therefore it is not sensitive to the degree of clustering. The data can be treated as if they really contained an I(d) process, i.e. differencing of degree d clears the evidence of the long memory property – the last two autocorrelation functions in the Figure 20 are decaying sufficiently fast after the same approach was applied to both processes.

The results from this part clearly support the third hypothesis. The estimated long memory parameter varies substantially based on the correlation between the directions of structural changes.

4.3.2 Manipulating the Size of Shifts

The simulations based on the data generating process described in Equation 19 yield very similar results to those run during the testing of the second hypothesis. As in all of the previous cases we start with the maximum number of lags equal to 5. The realized degree of clustering and the long memory parameter estimated on the data are presented in Figure 21 and the summary of the regression from Equation 22 can be found in Table 14. The estimate of the parameter δ_1 is equal to 0.067 which is a little lower than the result from the part 4.2.1, pattern with equally distributed probabilities. Also the increase in the degree of clustering over the 20 steps of the experiment is 39% and 42 % in the current setting and the one for the second hypothesis, respectively. The development of the mean estimate of the long memory parameter d seems to be less erratic over the 20 steps in the case with ARCH process in the breaks' magnitude. The Table 15 shows how the estimate of the coefficient δ_1 changes with higher maximum number of lags. We can say that in general, as in all previous cases except the one with sign-varying shifts, the estimate of the coefficient is increasing with increasing number of lags. Our idea why tampering with the magnitude of the shifts does not have any effect significantly different from the situation when we manipulate only the probabilities is that the clustered breaks still offset each other. We can think of it as that for each cluster of breaks it is randomly selected whether it consists of large or small shifts. The maximum number of lags only influences how long this cluster is – first, how many breaks will actually happen in a related sequence; second, how many past periods' size is determinant for the size of the current one. If we take a look at the Equation 19 we see that within the cluster the shifts have zero mean. Therefore their joint effect on the central process has a form of fluctuation within the periods in the cluster. On a bigger scale the effect is none, therefore the resulting estimated long memory parameter is influenced only by varying additional probabilities of break, exactly as it was during the testing of the second hypothesis.

The development of the estimate of coefficient δ_1 with increasing number of lags is not as smooth as it was in the case of the second hypothesis. We see a possible cause in some kind of shocks occurring in the behaviour of the final process. If we still perceive the problem as random sizes being assigned to each cluster we can imagine a situation when increasing the maximum lag above certain level causes two clusters to join into one and it mitigates the difference in size between them. That would imply that some levels of maximum lag (in relation to the sample length and the number of structural breaks occurring during the whole sample) can be connected to higher volatility in the sample mean and some other can be connected to a lower one, without any understandable regularity. However what we describe is only one possible explanation and we do not delve deeper into the problem.

In Figure 22 we present an example of shifts occurring during the whole sample with different levels of clustering and volatility clustering. From the graphs the clustering of the shifts of similar size is noticeable as well as the progressing clustering of breaks in time. The corresponding degrees of clustering and estimates of the long memory are in line with the logic described in the previous paragraph.

The experiment is repeated with the same alternative core processes as in the part 4.3.1. The results are presented in Table 16. There is no significant change in the estimated parameters and coefficients and still no higher smoothness of the development of the coefficient δ_1 estimate with increasing maximum number of lags.

We did not formulate a separate hypothesis for the manipulating of the shifts' size. However, it is still worth stating that we have found no evidence indicating that introduction of the volatility clustering into the data generating process has effect on the estimated long memory parameter.

4.4 Breaks' Number-Size Trade-off – Closing on the Real World Data

We say that a structural break occurs when one of the parameters of a model changes during the sample period. In our experiment we simplify such condition into changing the mean of the simulated data. As is stated in the Chapter 3 we incorporate such rules into the data generating processes that no matter how we tamper with the clustering of the breaks, their expected number within the sample remains the same. Such situation is necessary for the results to be comparable. So far we did not discuss how we set the total expected number of breaks.

In all of the realized simulations the total probability of break occurrence was set to 5%, which means that on average 50 breaks occur during the sample of 1000 observations. To put this in relation to a real world situation, such total probability of break corresponds to occurrence of 6 structural changes during ten years of monthly data which is a reasonable number. We do not examine the problem of how many breaks within a sample period can be still considered structural and from which number they should be incorporated into the estimated data generating process, therefore we need to provide more general results in terms of the total probability of breaks' occurrence. We simply want to compare the most significant results from the previous experiments to a situation with the same setting except for the parameter p_T.

In case that we lower the parameter and keep the additional probabilities of break occurrence after another break at the same level, these additional probabilities gain relatively more weight. Therefore we expect the measure of clustering to be more sensitive to higher maximum number of lags for lower values of the probability p_T. We will not consider any values higher than 5% in order to comply with the idea of rareness and irregularity of structural changes.

We select only the pattern with equally distributed additional probabilities of break occurrence (the data generating process is described in the part 3.2, the results from the previous testing based on this setting are described in the part 4.2.1). We do not expect that changing the number of breaks would have different results for the other two patterns. The lowest expected number of breaks we want to test is 10 within the whole sample, i.e. the parameter p_T equal to 1%. Since the weight of additional probabilities of break is relatively higher with respect to the independent probability of break than in the case when the parameter p_T was higher, we expect the degree of clustering to increase more rapidly with further steps of the experiment. We can imagine the change of the parameter p_T as a situation in which we already know when some number of breaks happens during the sample and in which we subsequently remove some of those breaks. Given our setting it is more probable that removed are those that stand alone and do not start a cluster. Therefore the median distance between breaks increases relatively more in comparison to the maximum possible mean distance. If we keep in mind the formula for our degree of clustering described in Equation 4 we see that with such changes its resulting value is higher. It is sufficient to examine this impact of changing the number of breaks on the degree of clustering in terms of the overall change of the measure, i.e. the difference between its maximum and minimum values over the 20 steps of the experiment. Using the mean degree of clustering of the usual 200 repetitions eliminates the effect of random elements. In order to get more detailed results capturing the role of the parameter p_T in the model we perform the whole experiment with values ranging from 1% to 4%.

The Table 17 shows what impact the change of the parameter p_T has on the estimate of the coefficient δ_1 and its effect on the overall increase of the mean degree of clustering over the 20 steps. All the estimates of the coefficient δ_1, which measures the relationship between the degree of clustering and the estimated long memory parameter, are significant on 1% level. From the results we see that our expectations were correct. Since the estimates of the coefficient are lower for the majority of maximum lags' values even though the increase of the degree of clustering is higher we can see that the effect of clustering on the estimated long memory parameter is weaker. As an example we present the Figure 23 showing the plot of the mean estimated parameter d over the 20 steps for the case with p_T equal to 1% and with 10 maximum lags. The development is clearly more volatile than for higher total probability of breaks' occurrence.

The results for all different values of parameter p_T are all in line with the same logic. The described effects are weaker for higher values of p_T and stronger for lower values. Even though the relationship between the clustering and the long memory is not so distinctive for the lower values it would still be sufficient for us to draw the same conclusions as before (as in the situation with 5% total probability of break).

The conclusion of the experiment with the total probability of structural change provides insight which is seemingly not favourable for our previous results. If it really was capturing the whole mechanism of translation from breaks' clustering to long memory it would mean that all our previous results are valid only for the specific case with the total probability of break equal to 5%. However we encountered a similar problem with the opposite effect earlier on – in the part 4.2.1 we noticed that the size of breaks is pressing the estimated long memory parameter against its maximum value. We solved it by changing the rule for the variance of the breaks-generating process. That is where the idea of an offsetting relationship between the size of breaks and their number comes from. The rule which sets the variance of the breaks' process equal to one fifth of the variance of the innovations to the core process is as artificial as was the total probability of breaks' occurrence; therefore we can manipulate it freely to see if we can balance out the lower number of breaks. We take the situation with the total probability of break equal to 1% and change the ratio of the two variances. The minimum value for the ratio is one fifth, i.e. 0.2, and we scale it up to 0.5.

The Table 18 shows the results in the same structure as did the Table 17, only for the case of changing the ratio between the variance of the breaks' distribution and the variance of the core process, σ_B^2/σ_C^2. Since we do not manipulate the number of breaks or their timing (p_T = 1%), the overall increase in the degree of clustering remains the same for equal number of maximum lags. However the estimated coefficient δ_1 increases significantly. That must mean that the estimate of the long memory parameter d is higher as well since the degree of clustering is still increasing the same. With the ratio of variances equal to 0.5 we receive even higher estimates of the coefficient δ_1 than we had with the parameter p_t equal to 5%.

The results yielded by this experiment imply that even though the total number of breaks within the sample is questionable the number alone is not determining for the strength of the relationship between the degree of clustering and the estimate of parameter d. The mix of breaks' number and size is the key condition – the dominance of the structural changes over the core process can be reached by increasing either of them. Therefore setting the parameter p_T to some value does not limit the conclusions drawn from the results.

5 Conclusion

Forecasting is often complicated by the occurrence of structural breaks during the data sample. It seems to be especially problematic when the data generating process of the original data is from the ARFIMA(p,d,q) family. For accurate forecasting all parameters of the data generating process have to be estimated as precisely as possible, including the parameter of fractional integration, d. According to current research the estimate of this parameter d can be easily contaminated by the long memory – slowly decaying autocorrelations – brought by structural breaks. A possible solution to this problem is to understand how exactly the estimate reacts to specific periods of structural changes and consequently correct the bias. In this thesis we have examined the effect of structural breaks' clustering on the estimate of the parameter d. Using Monte Carlo simulations we have shown that different patterns of structural breaks cause different bias. Its size is affected by the intensity of clustering, the length of the clusters and the signs of structural changes within them. We have shown that correlation between signs of the structural changes has the strongest influence on the estimate. Clustering of shifts to mean with positively correlated signs magnifies the overestimate of the parameter d while negative correlation between signs reduces it substantially. Our simulations also indicate that volatility clustering during the transition periods has no effect on the estimate of the long memory parameter. Our research can be followed by developing estimators of the long memory parameter that take into consideration the properties of structural breaks' distribution within the sample while our measure of clustering can be used to quantify its intensity. Together with better methods for breaks' identification, the findings presented in this thesis open way to increasing the performance of forecasting subject to structural breaks and fractional integration.

Bibliography

Beran, Y. Feng, S. Ghosh & R. Kulik (2013): Long-Memory Processes: Probabilistic Properties and Statistical Methods. Springer Science & Business Media.

Beran, J. (1994): Statistics for Long Memory Processes. CRC Press.

Robinson, P. M. (1995): "Log-periodogram Regression of Time Series with Long Range Dependence." The Annals of Statistics: pp. 1048-1072.

Shimotsu, K. et al. (2005): "Exact Local Whittle Estimation of Fractional Integration." The Annals of Statistics: pp. 1890-1933.

Granger, C. W. J. & N. Hyung (2004): "Occasional Structural Breaks and Long Memory with an Application to the S&P 500 Absolute Stock Returns." Journal of Empirical Finance 11.3: pp. 399-421.

Wang, C. S. H., L. Bauwens & C. Hsiao (2013): "Forecasting a Long Memory Process Subject to Structural Breaks." Journal of Econometrics 177.2: pp. 171-184.

Satchell, S. & J. Knight (2011): Forecasting Volatility in the Financial Markets. Butterworth-Heinemann.

Poon, S. H. & C. W. J. Granger (2003): "Forecasting Volatility in Financial Markets: A Review." Journal of Economic Literature 41.2: pp. 478-539.

Hansen, P. R. & A. Lunde (2005): "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?." Journal of Applied Econometrics 20.7: pp. 873-889.

Shepard, N. (1996): "Statistical Aspects of ARCH and Stochastic Volatility." Monographs on Statistics and Applied Probability 65: pp. 1-68.

Reisen, V. A. & S. Lopes (1999): "Some Simulations and Applications of Forecasting Long-Memory Time-Series Models." Journal of Statistical Planning and Inference 80.1: pp. 269-287.

Abraham, B. & V. A. Reisen (1998): Prediction of Long Memory Time Series Models: A Simulation Study and an Application. University of Waterloo.

Bai, J. & P. Perron (2003): "Computation and Analysis of Multiple Structural Change Models." Journal of Applied Econometrics 18.1: pp. 1-22.

Perron, P. & X. Zhu (2005): "Structural Breaks with Deterministic and Stochastic Trends." Journal of Econometrics 129.1: pp. 65-119.

Busetti, F. (2002): "Testing for (Common) Stochastic Trends in the Presence of Structural Breaks." Journal of Forecasting 21.2: pp. 81-105.

Clements, M. & D. Hendry (1998): Forecasting Economic Time Series. Cambridge University Press.

Granger, C. W. J. (1966): "The typical spectral shape of an economic variable." Econometrica: Journal of the Econometric Society: pp. 150-161.

Xu, J. & P. Perron (2014): "Forecasting return volatility: Level shifts with varying jump probability and mean reversion." International Journal of Forecasting 30.3: pp. 449-463.

Lu, Y. K. & P. Perron (2010): "Modeling and forecasting stock return volatility using a random level shift model." Journal of Empirical Finance 17.1: pp. 138-156.

Hidalgo, J. & P. M. Robinson (1996): "Testing for structural change in a long-memory environment." Journal of Econometrics 70.1: pp. 159-174.

Beine, M. & S. Laurent (2000): "Structural change and long memory in volatility: new evidence from daily exchange rates." Econometric Society World Congress.

Krämer, W. & P. Sibbertsen (2000): "Testing for structural change in the presence of long memory." Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund.

Granger, C. W. J. & R. Joyeux (1980): "An introduction to long-memory time series models and fractional differencing." Journal of time series analysis 1.1: pp. 15-29.

Hosking, J. R. M. (1981): "Fractional differencing." Biometrika 68.1: pp. 165-176.

Appendix A: Figures

Author's own computations are the source of all figures.

Figure 1: Bias-variance trade-off based on truncation parameter m

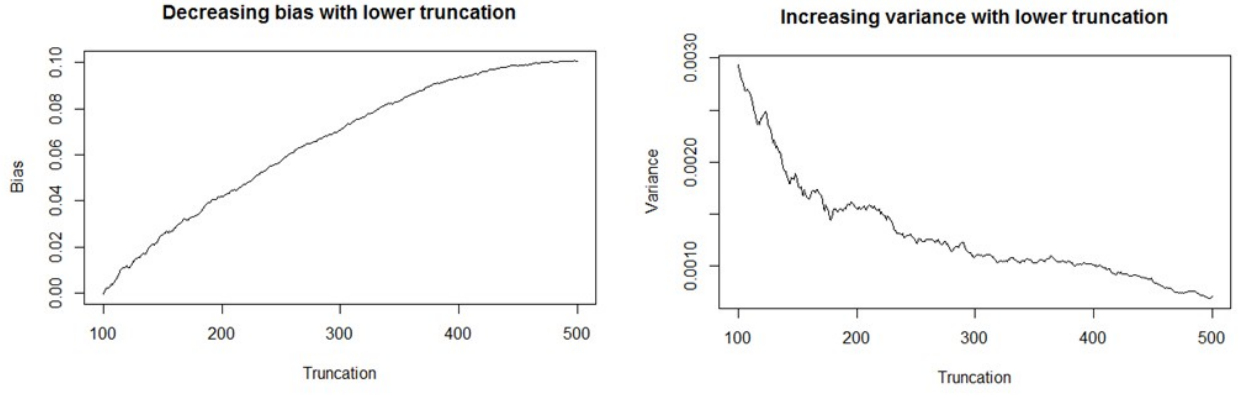
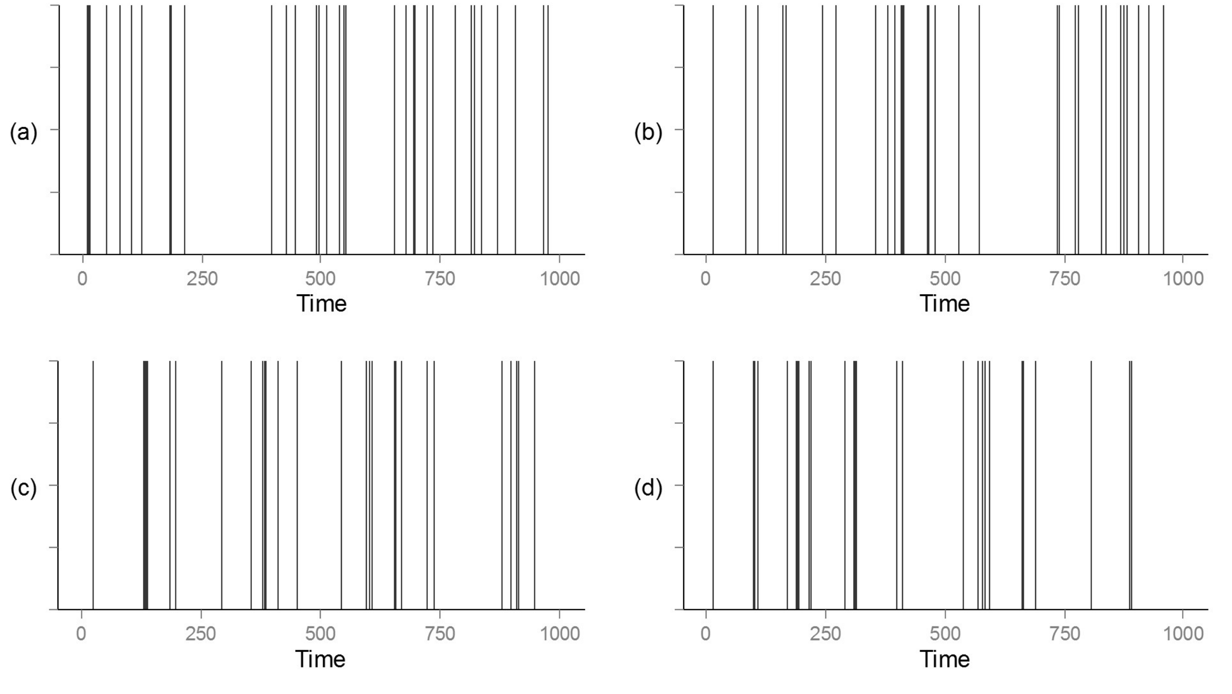
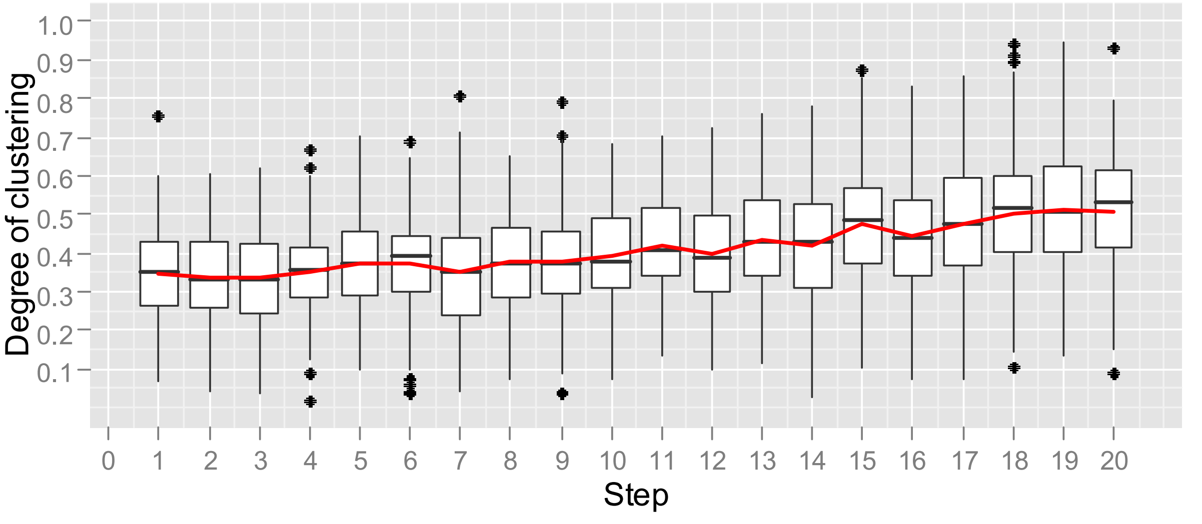


Figure 2: Distribution of breaks



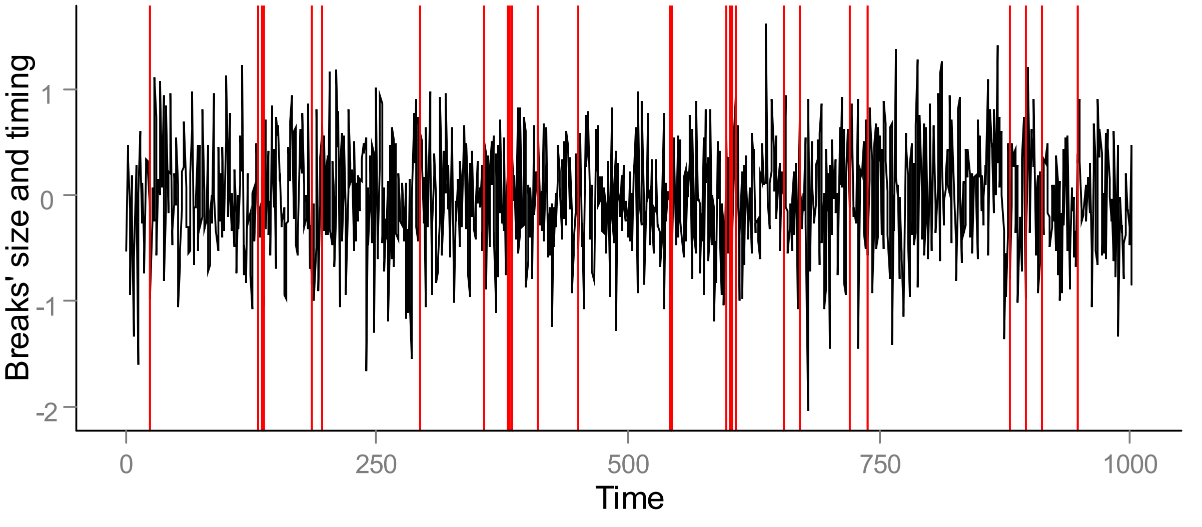
An example of distribution of breaks for the testing of the first hypothesis, p_1 equal to 0.005 (a), 0.035 (b), 0.070 (c) and 0.100 (d)

Figure 3: Degree of clustering, first hypothesis



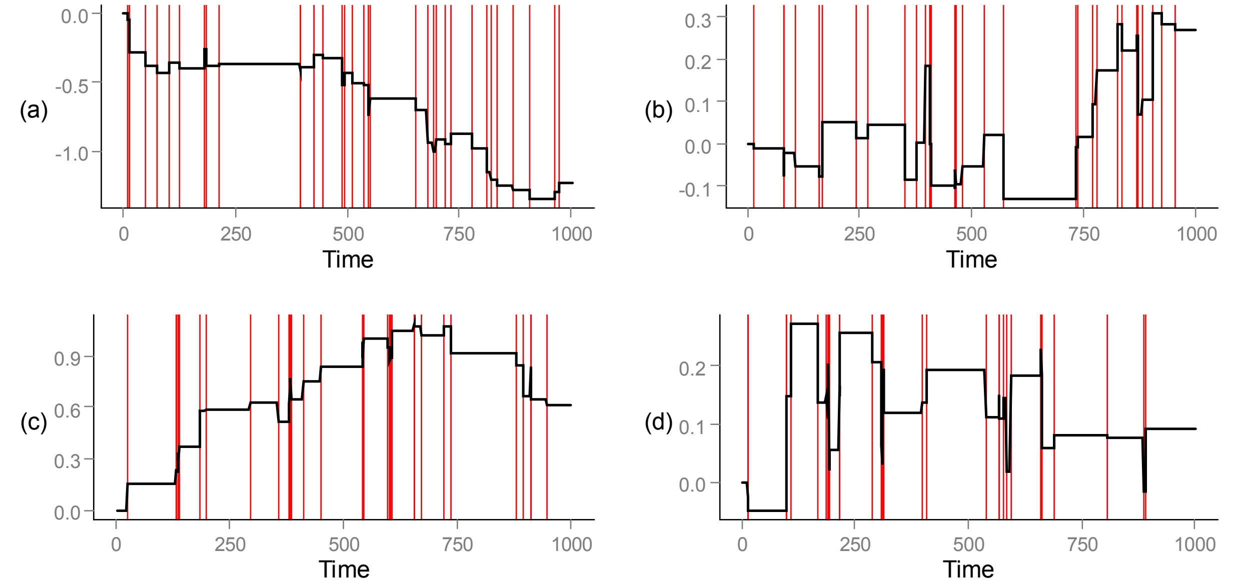
The calculated degree of clustering of the simulated data; the line represents the mean value.

Figure 4: Breaks' size and timing



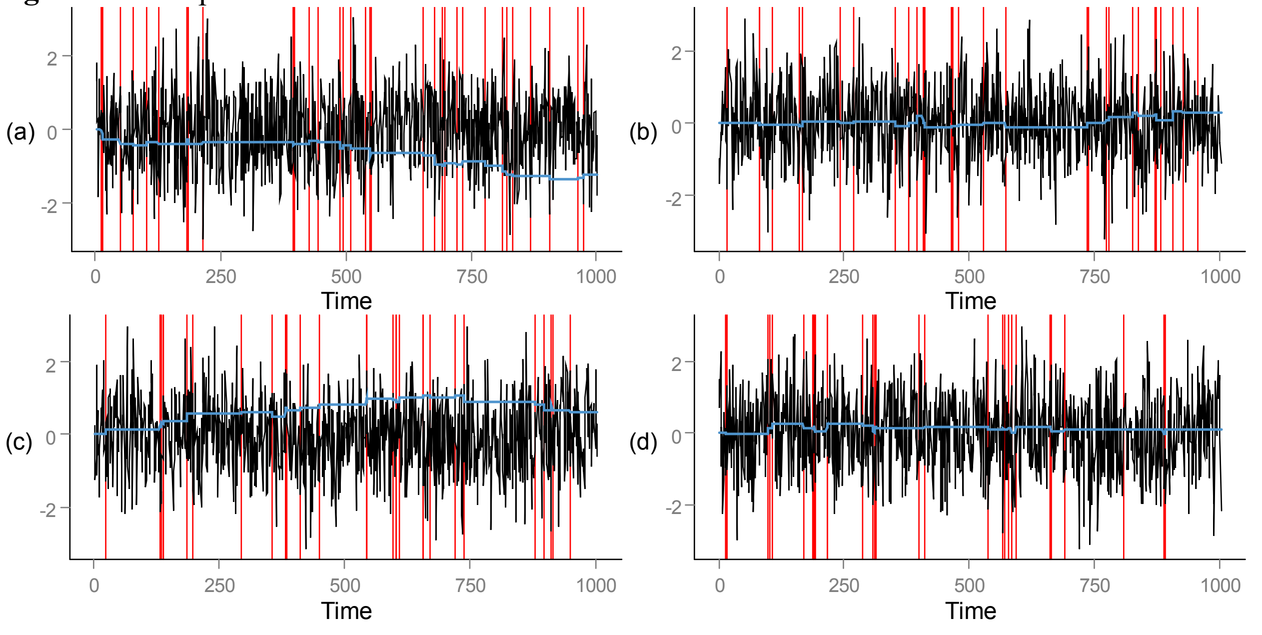
A representation of how the vector of shifts to mean is generated

Figure 5: Cumulative shift to mean



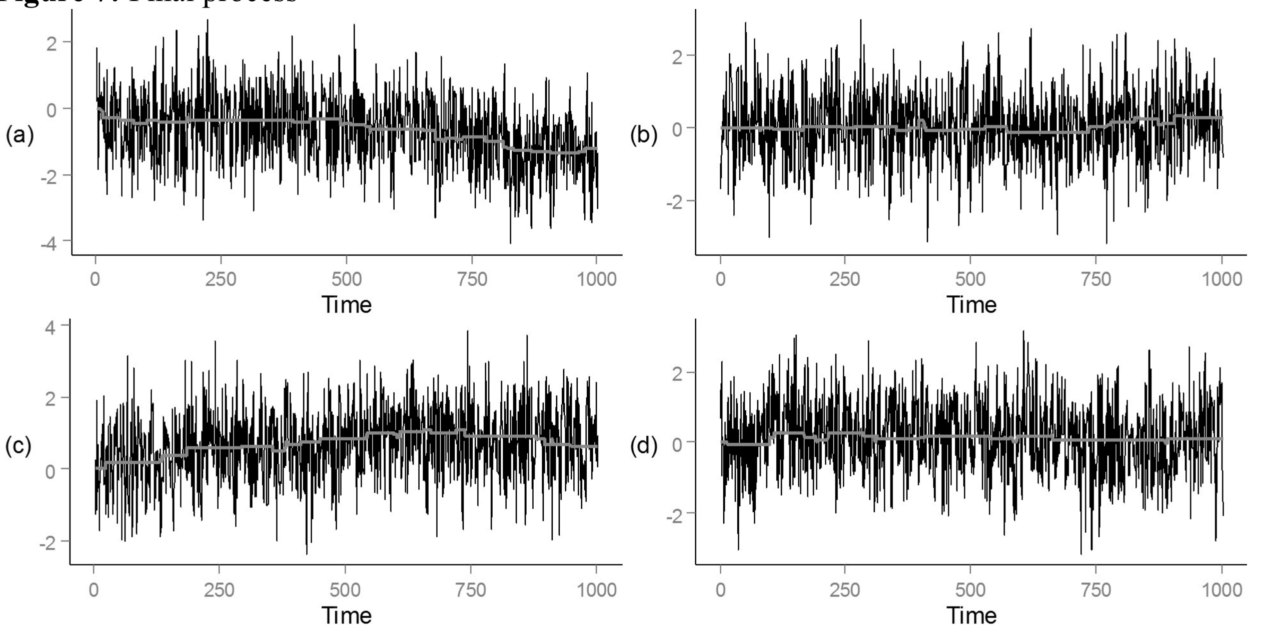
An example of the cumulative shift to mean with marked periods in which the breaks occur. p_1 is equal to 0.005 (a), 0.0350 (b), 0.0700 (c) and 0.1000 (d).

Figure 6: Core process and the cumulative shift to mean



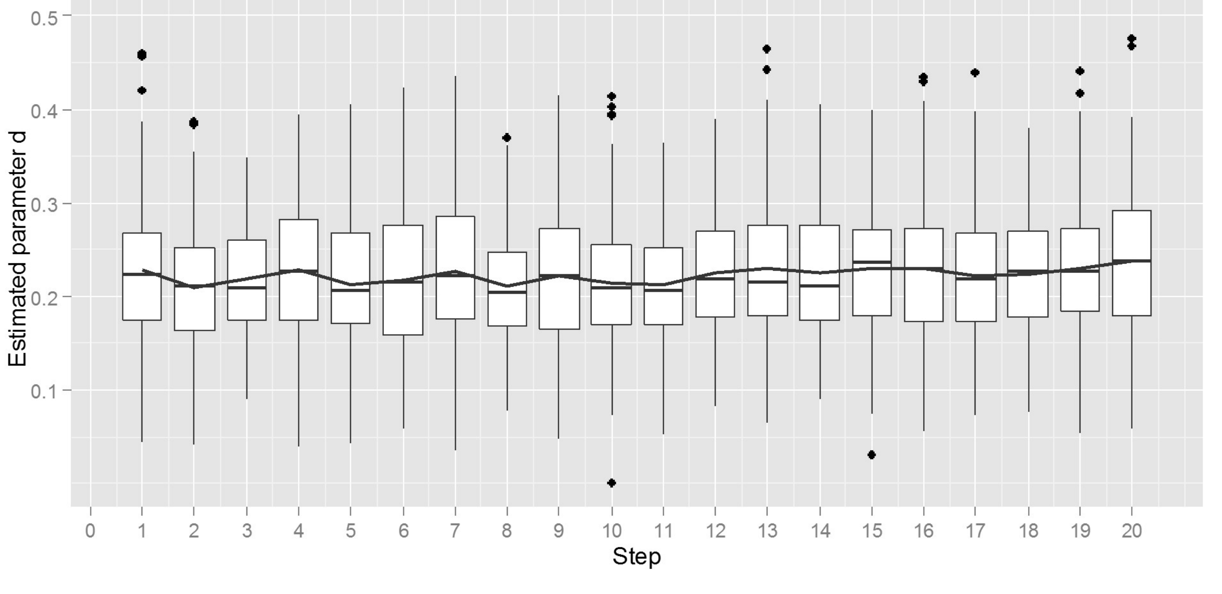
An example of the core process with marked timing of breaks. The thick line represents the cumulated shift to mean. p_1 is equal to 0.005 (a), 0.0350 (b), 0.0700 (c) and 0.1000 (d).

Figure 7: Final process



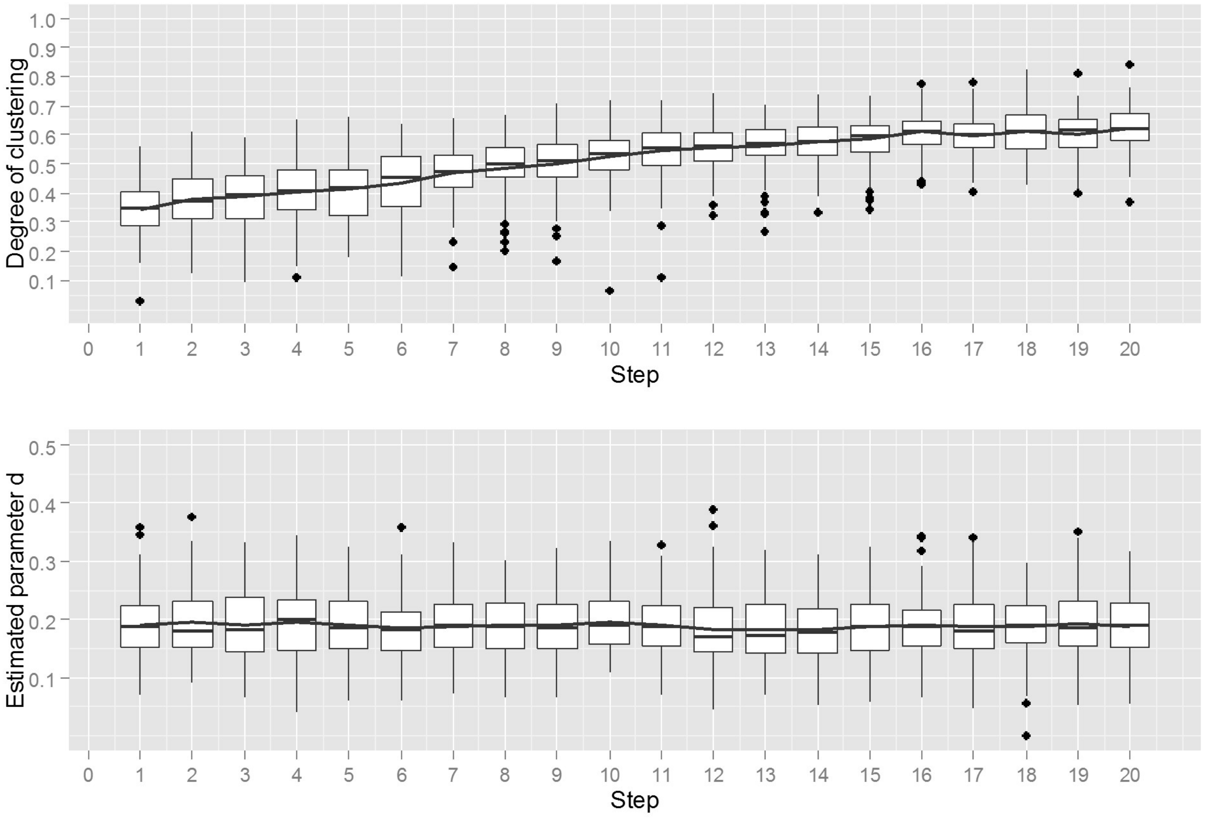
An example of the final generated process. The thick line represents the shift to mean which was added to the core process. p_1 is equal to 0.005 (a), 0.0350 (b), 0.0700 (c) and 0.1000 (d).

Figure 8: Estimated long memory parameter, first hypothesis



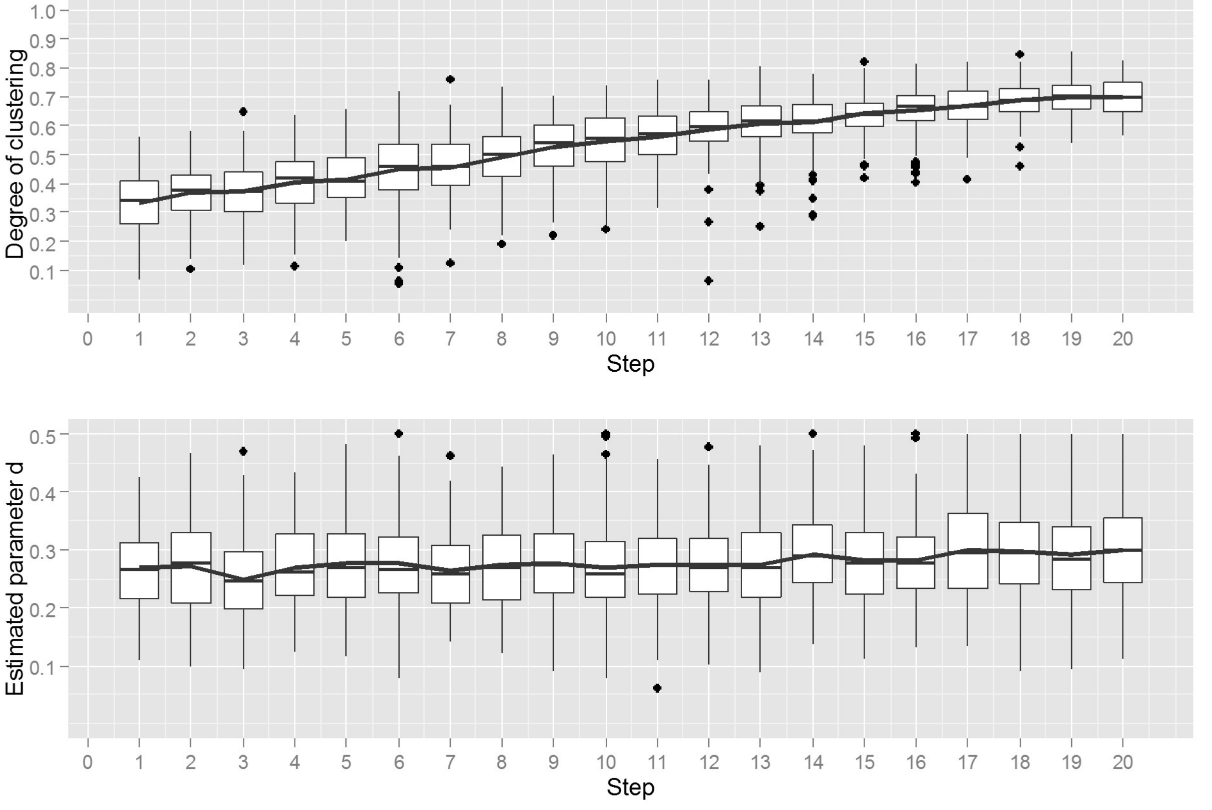
The estimates of the long memory parameter. The line represents the mean value.

Figure 9: First hypothesis, alternative core process, insignificant relationship



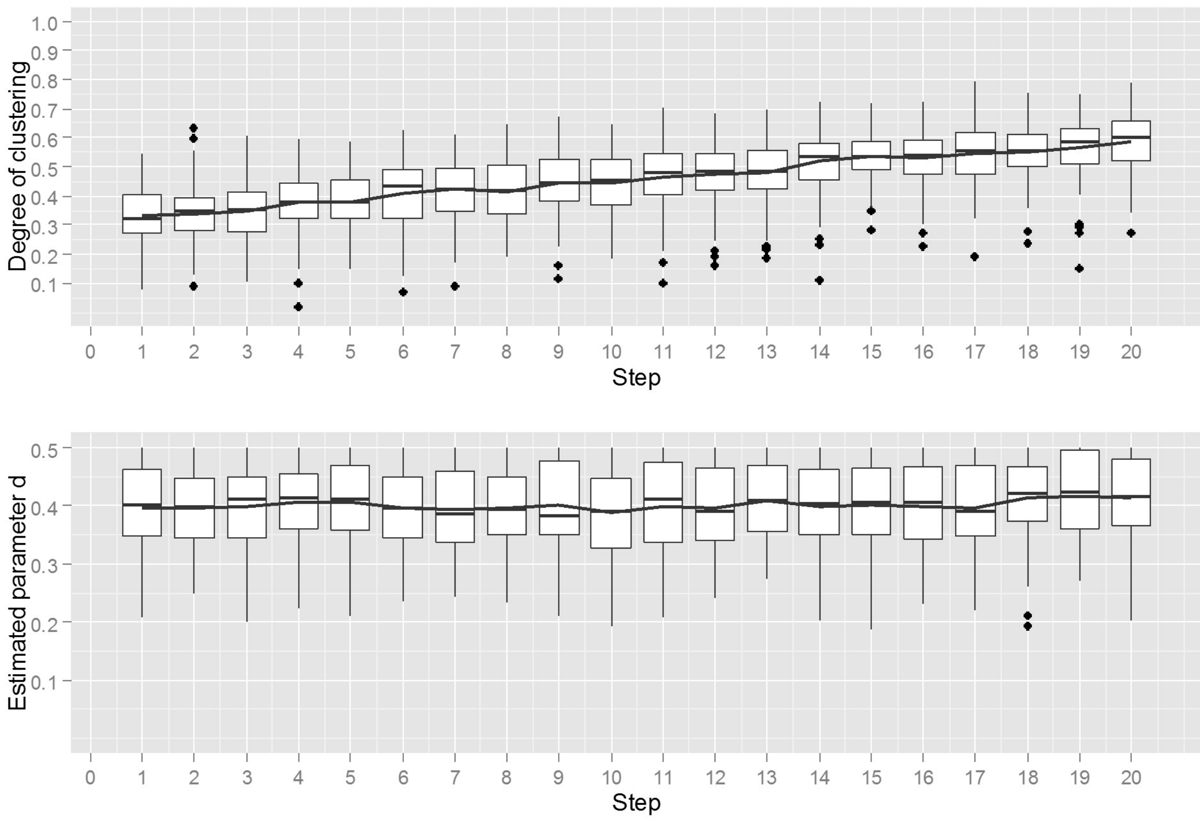
The degree of clustering and the estimates of d. The line represents the mean value.

Figure 10: First hypothesis, alternative core process, significant relationship



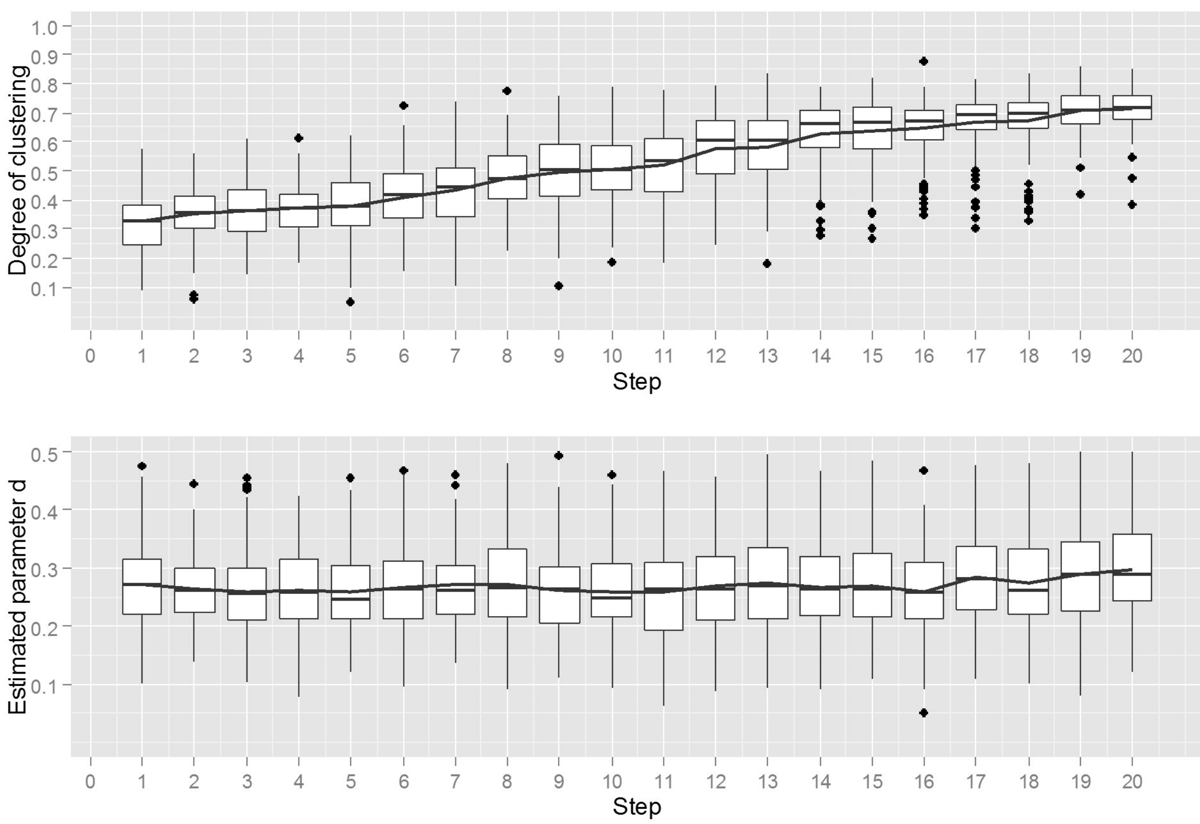
The degree of clustering and the estimates of d. The line represents the mean value.

Figure 11: Second hypothesis, equal probabilities



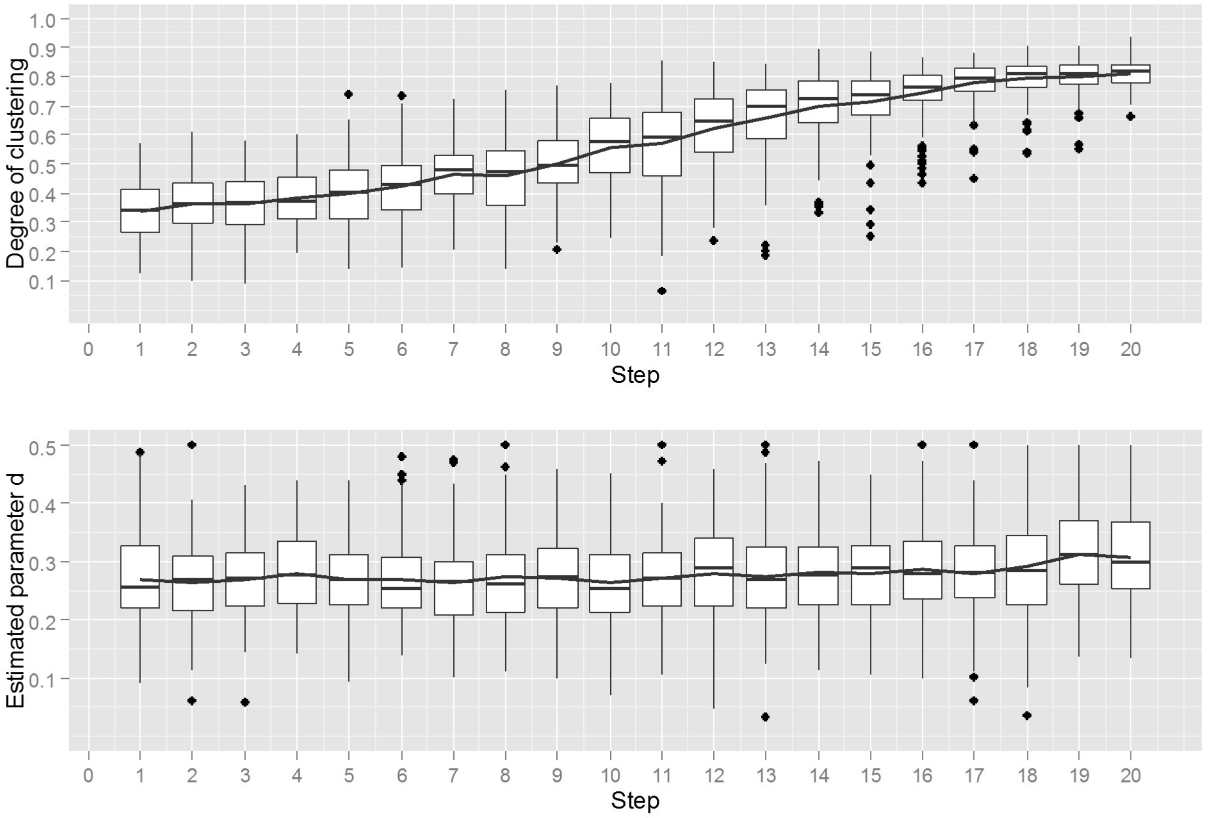
The degree of clustering and the estimates of d. The line represents the mean value.

Figure 12: Second hypothesis, progressive probabilities



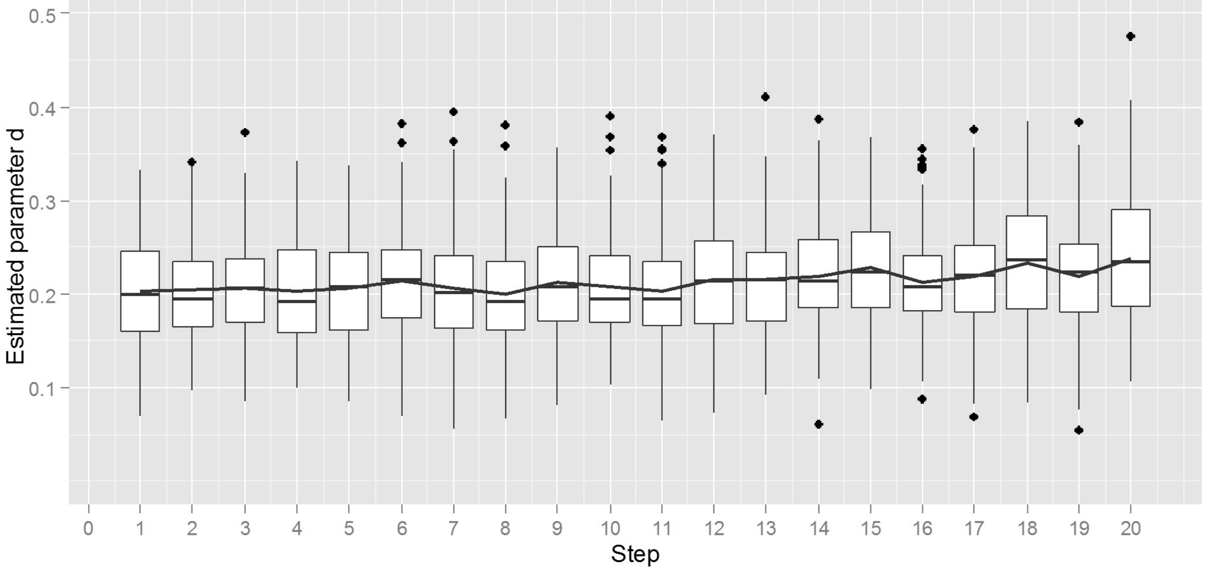
The degree of clustering and the estimates of d. The line represents the mean value.

Figure 13: Second hypothesis, increased importance of earlier periods



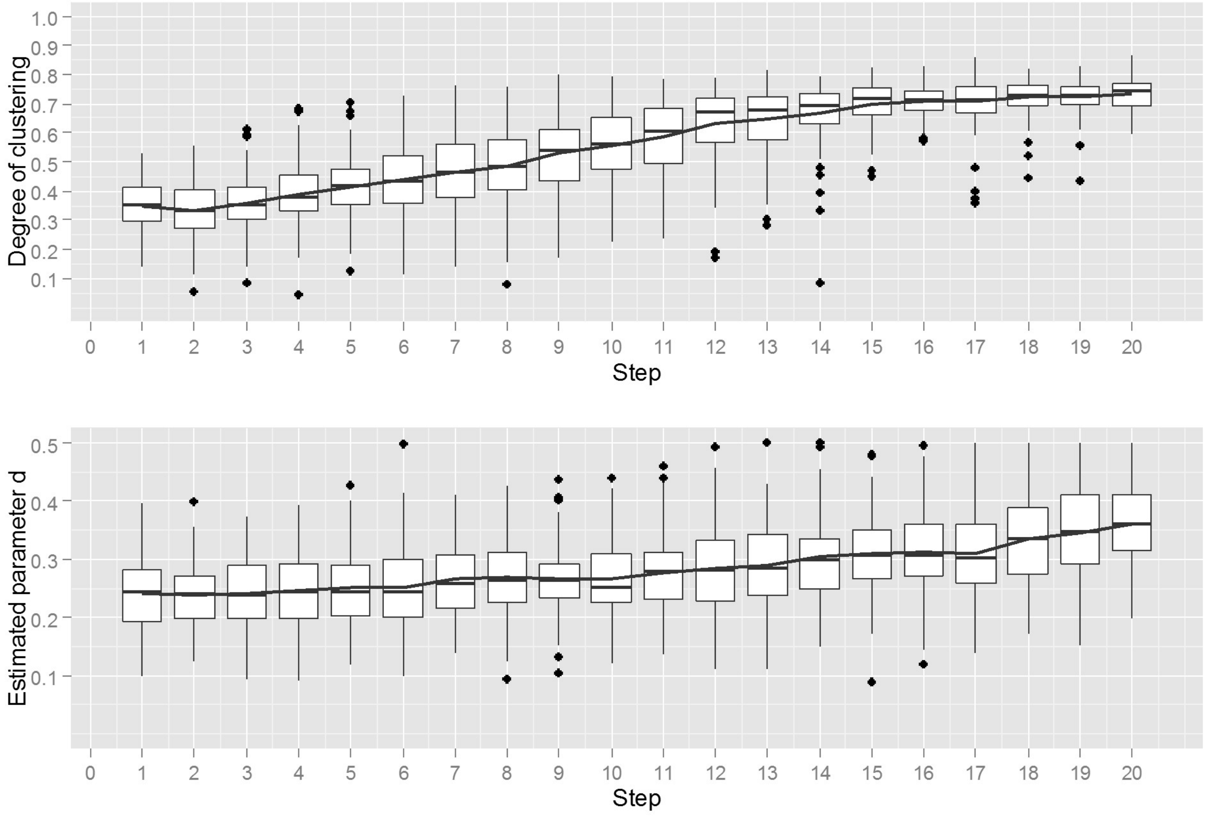
The degree of clustering and the estimates of d. The line represents the mean value.

Figure 14: An example of estimates with different truncation parameter m



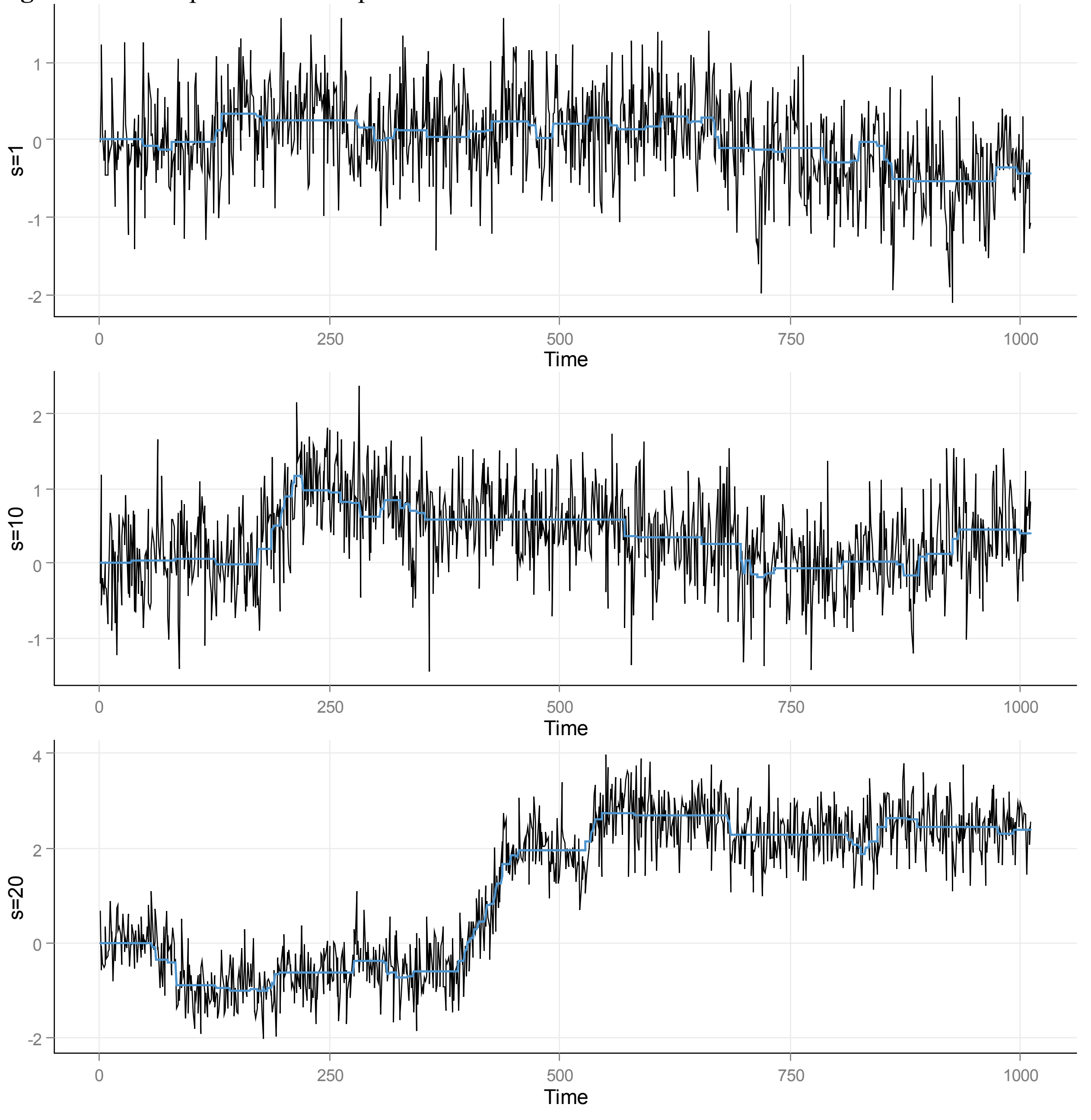
The estimates of the long memory parameter. The line represents the mean value.

Figure 15: Third hypothesis, same-sign shifts



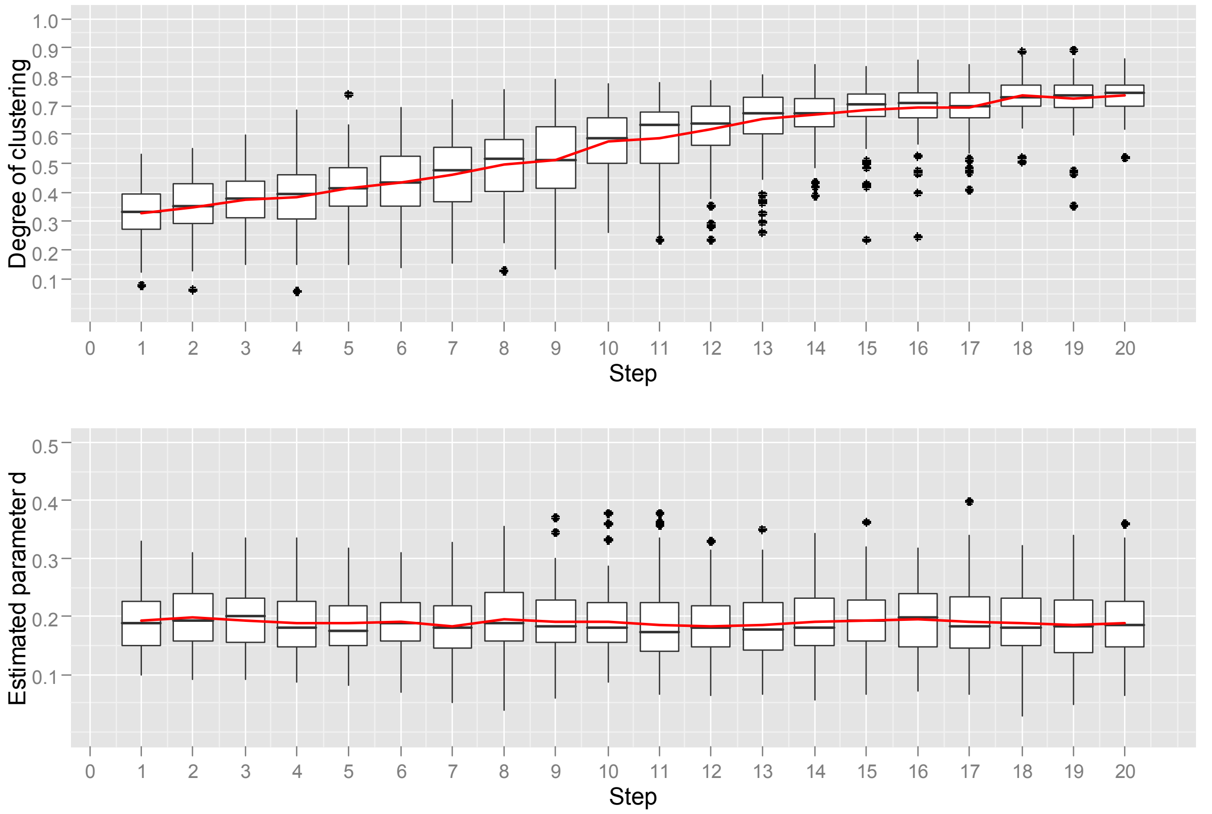
The degree of clustering and the estimates of d. The line represents the mean value.

Figure 16: Example of the final processes and cumulative shifts to mean



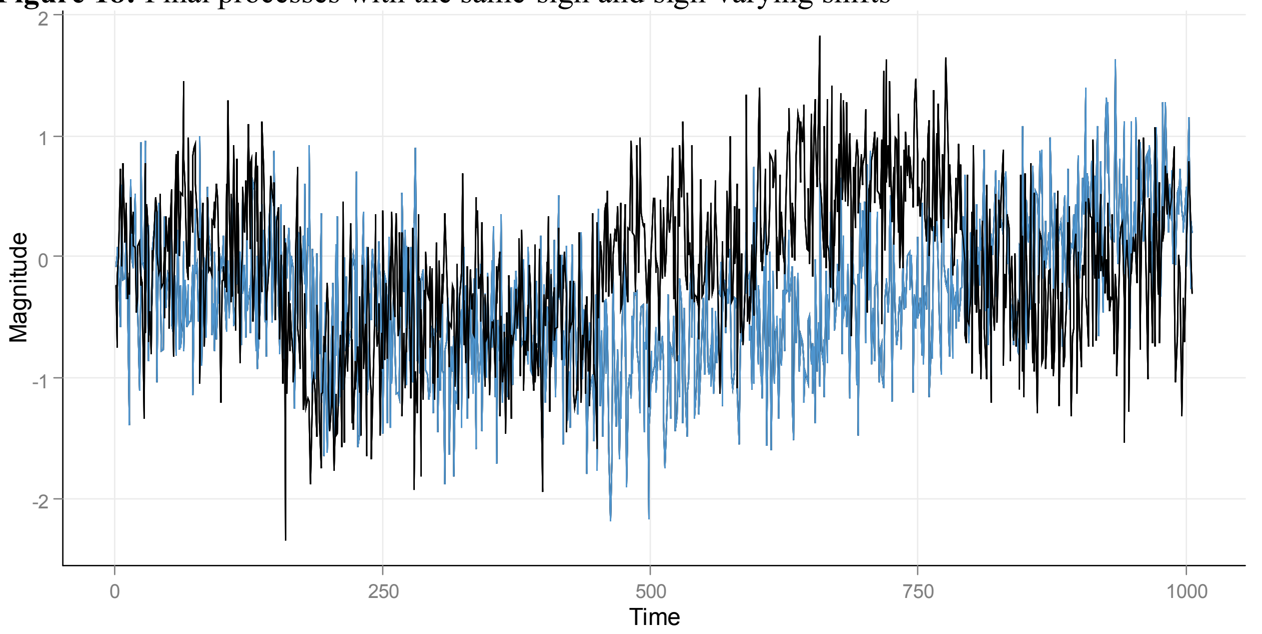
An example of the final process with different degree of clustering. The thick line represents the mean value.

Figure 17: Third hypothesis, sign-varying shifts



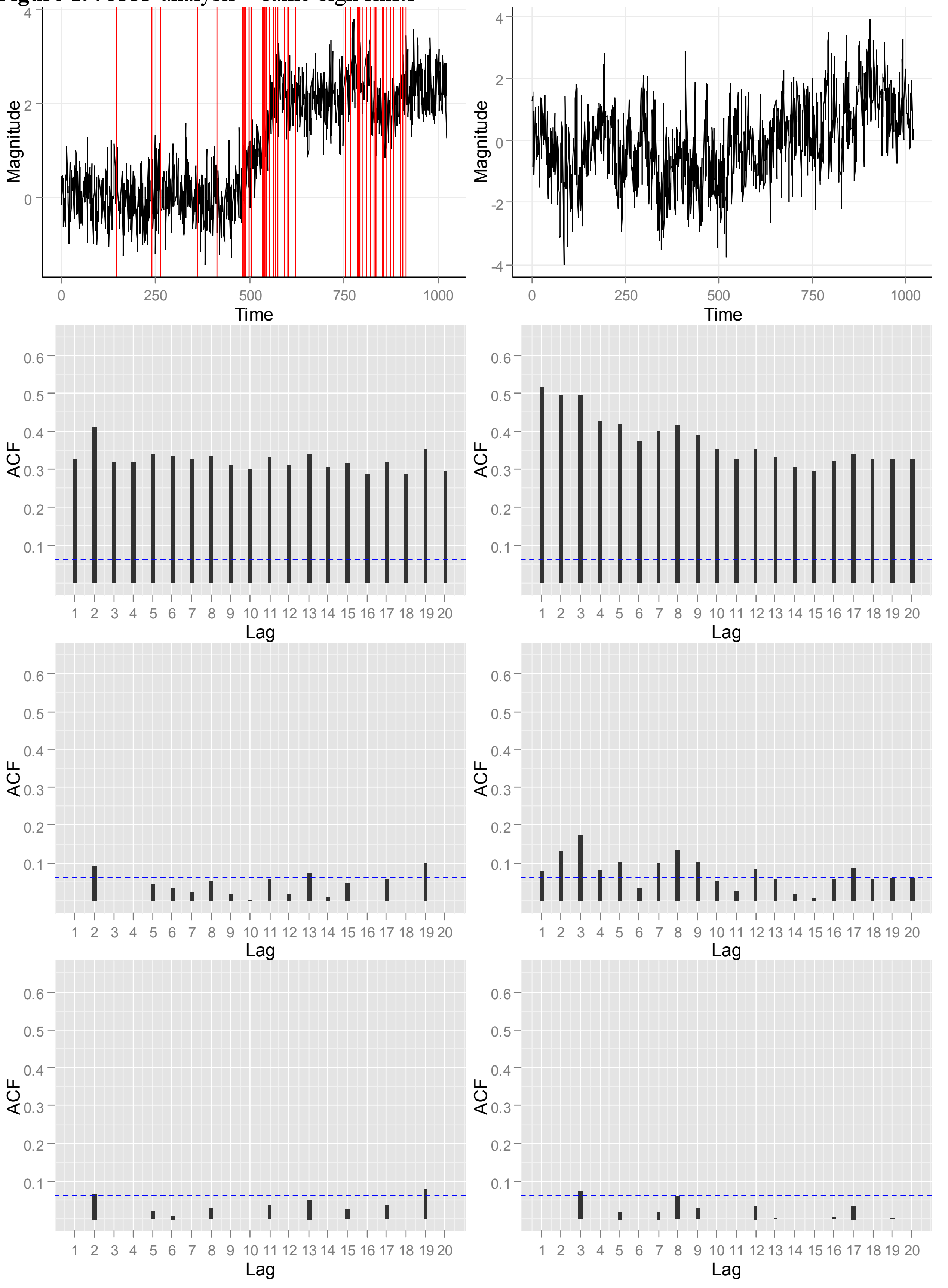
The degree of clustering and the estimates of d. The line represents the mean value.

Figure 18: Final processes with the same-sign and sign-varying shifts



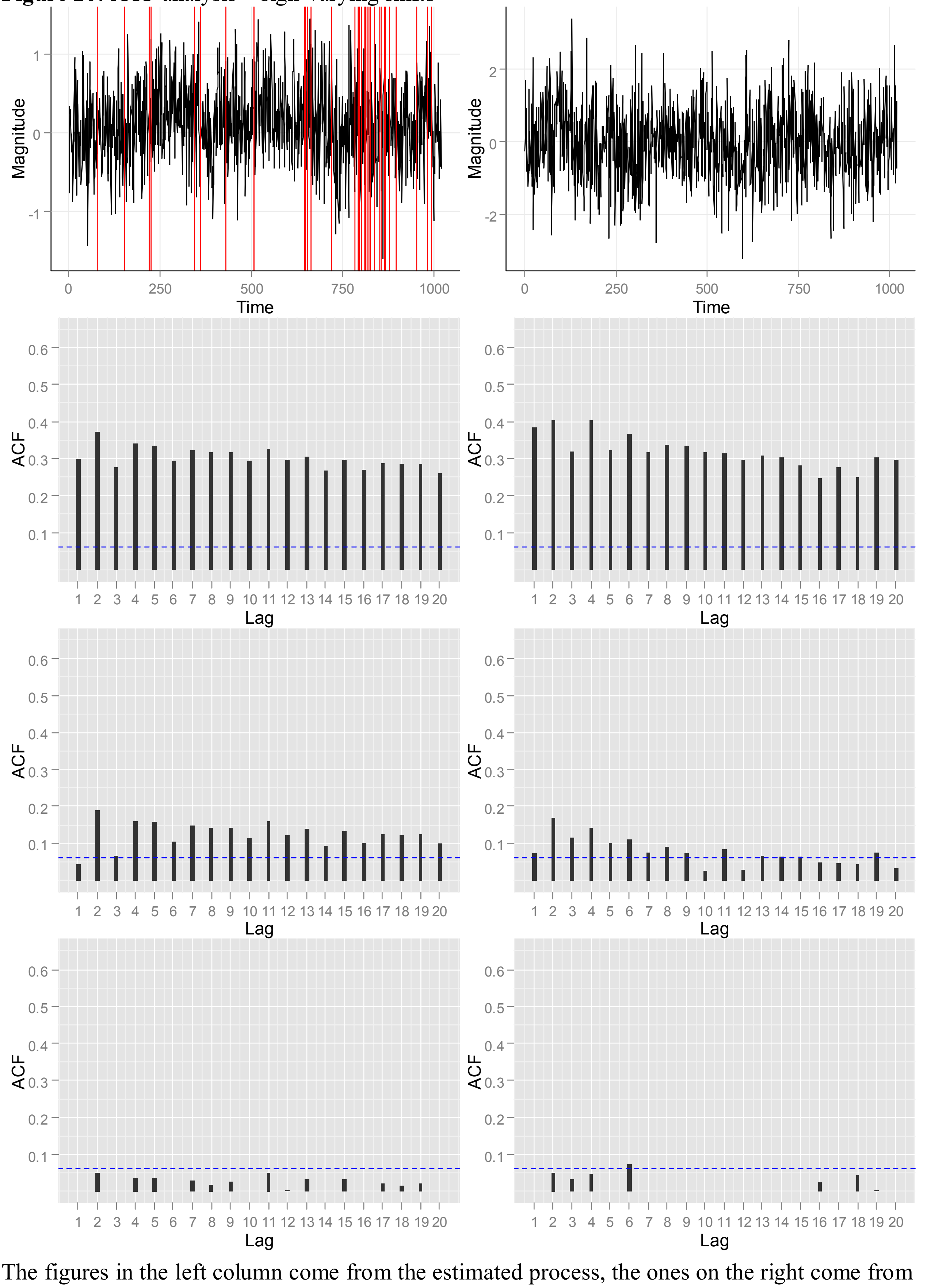
An example of the final process with same-sign (dark) and sign-varying (bright) shifts

Figure 19: ACF analysis – same-sign shifts



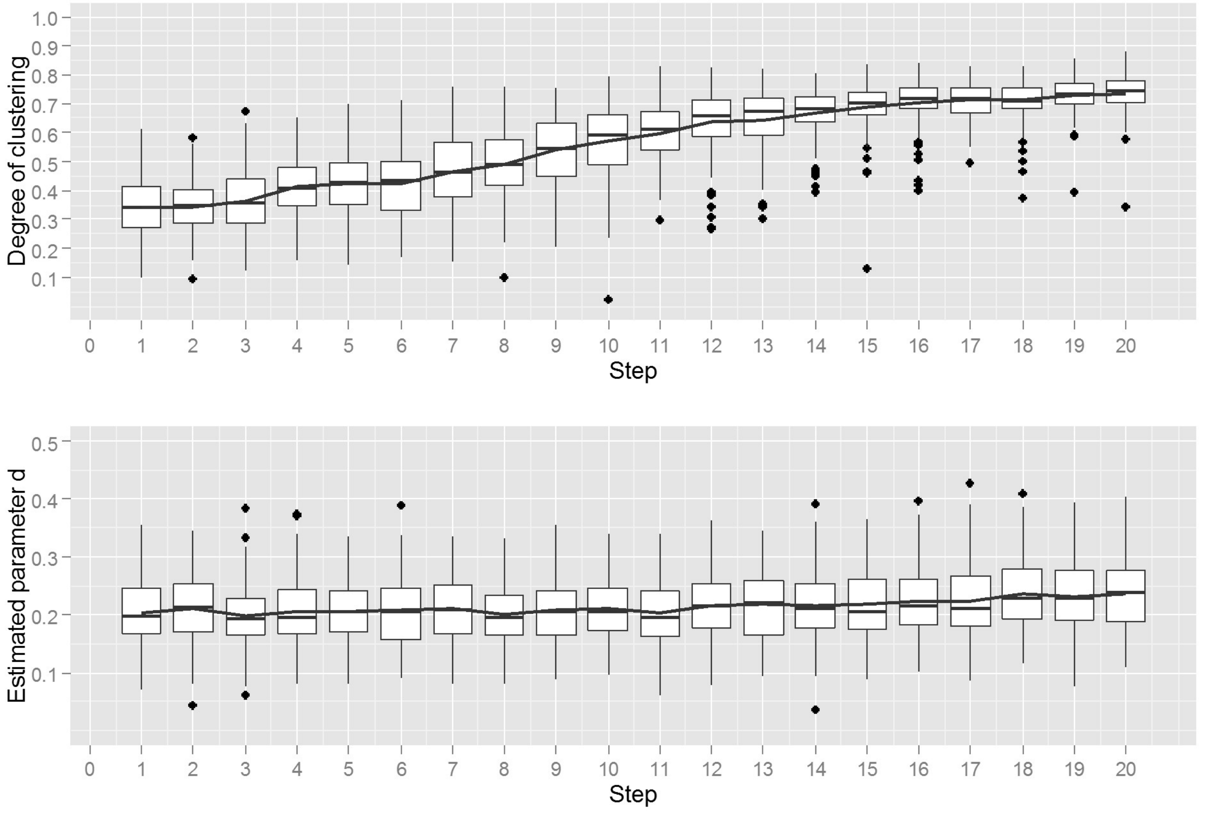
The figures in the left column come from the estimated process, the ones on the right come from the ARFIMA. The second line shows the ACF of the processes, third and fourth show the ACF of the processes differentiated by 75% and 100% of the estimated value of d, respectively.

Figure 20: ACF analysis – sign-varying shifts



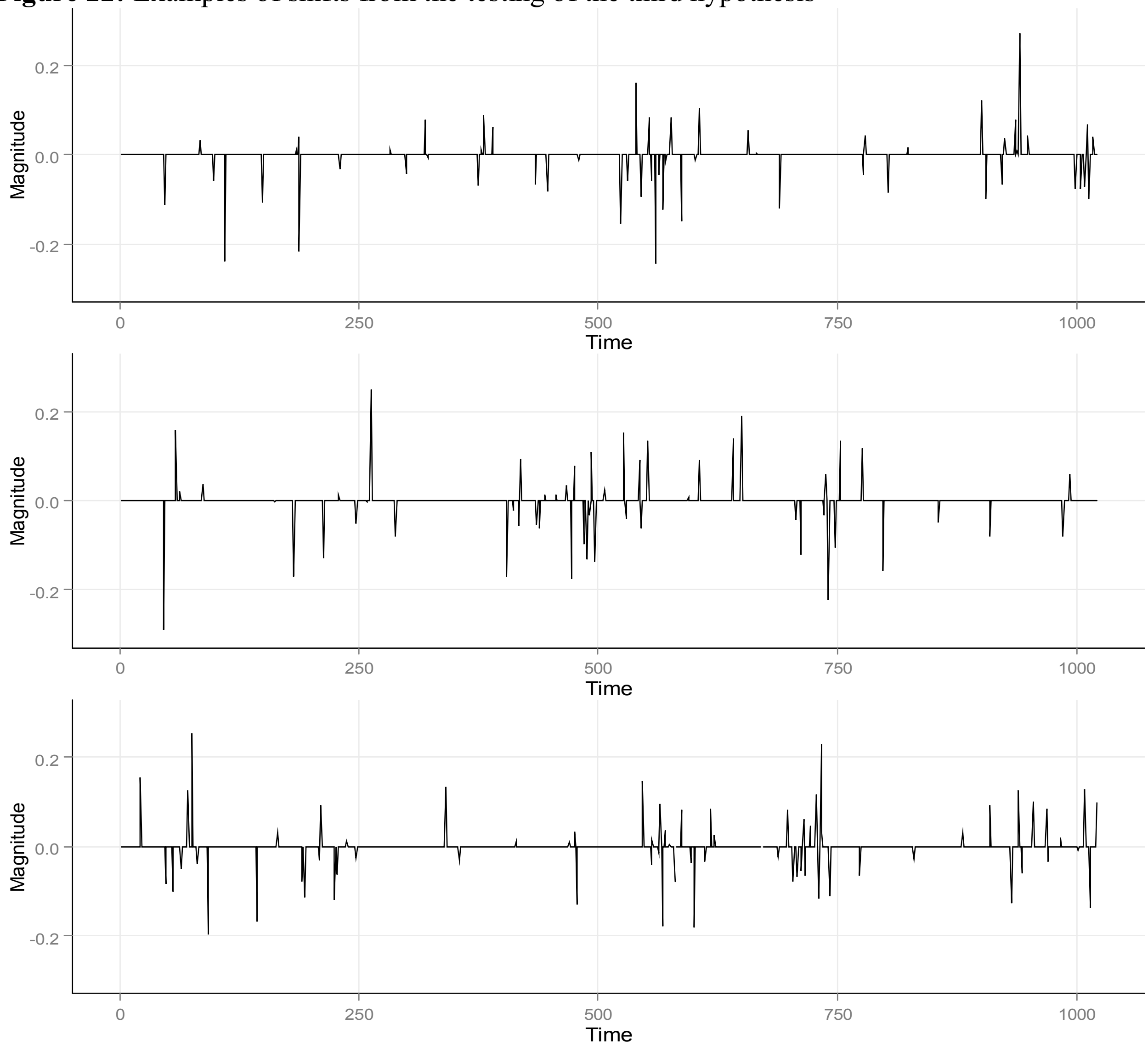
The figures in the left column come from the estimated process, the ones on the right come from the ARFIMA. The second line shows the ACF of the processes, third and fourth show the ACF of the processes differentiated by 75% and 100% of the estimated value of d, respectively.

Figure 21: Third hypothesis, manipulating the size of shifts



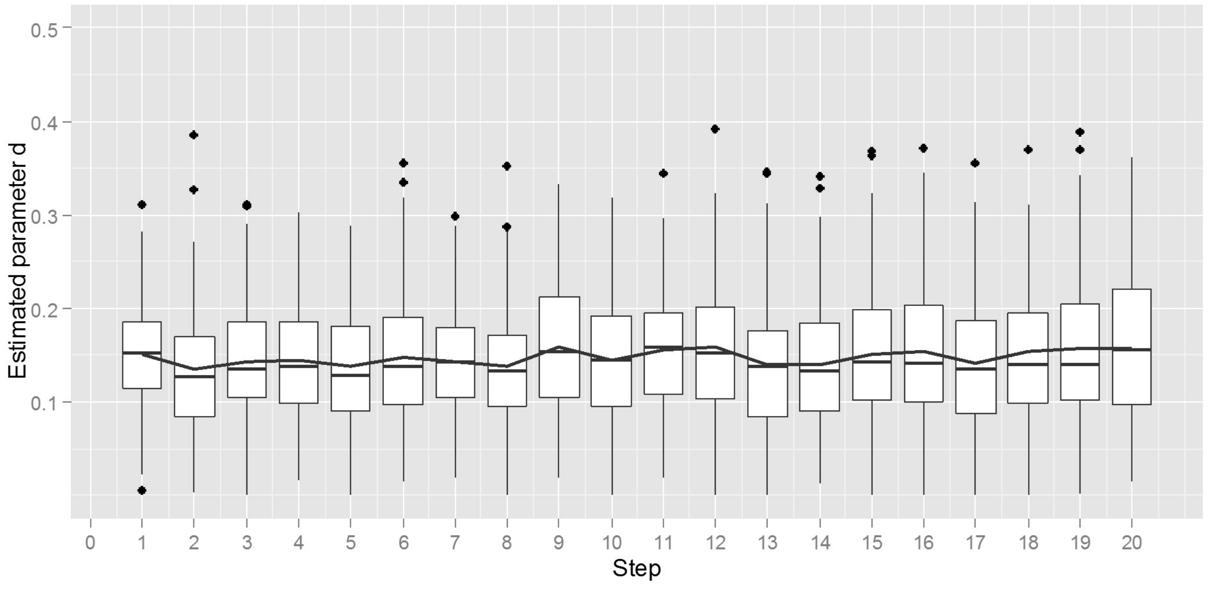
The degree of clustering and the estimates of d. The line represents the mean value.

Figure 22: Examples of shifts from the testing of the third hypothesis



The timing and size of shifts representing the volatility clustering. Plots from the step 1, 10 and 20, respectively.

Figure 23: The estimate of the parameter d of process with lower total probability of breaks' occurrence



The estimates of the long memory parameter. The line represents the mean value.

Appendix B: Tables

Table 1: First hypothesis – regression summary

	Estimate	Std. error	t value	P(>|t|)	
δ_0	0.192	0.011	17.331	$1.12*e^{-12}$	***
δ_1	0.071	0.027	2.635	0.017	*
R^2	0.278	Significance codes: *** 0.001, ** 0.01, * 0.05, + 0.1			

Table 2: Second hypothesis, equal probabilities – regression summary

	Estimate	Std. error	t value	P(>|t|)	
δ_0	0.237	0.008	27.603	$3.48*e^{-16}$	***
δ_1	0.074	0.015	5.005	$9.19*e^{-05}$	***
R^2	0.581	Significance codes: *** 0.001, ** 0.01, * 0.05, + 0.1			

Table 3: Second hypothesis, equal probabilities – different values of k

k	$\widehat{\delta_1}$	Std. error	P(>|t|)		R^2
5	0.074	0.015	$9.19*e^{-05}$	***	0.581
6	0.059	0.010	$3.88*e^{-05}$	***	0.619
7	0.081	0.015	$3.94*e^{-05}$	***	0.618
8	0.085	0.016	$4.14*e^{-05}$	***	0.616
9	0.101	0.018	$1.72*e^{-05}$	***	0.651
10	0.099	0.021	$2.11*e^{-04}$	***	0.543
15	0.153	0.022	$1.45*e^{-06}$	***	0.733
20	0.174	0.032	$3.29*e^{-05}$	***	0.626
		Significance codes: *** 0.001, ** 0.01, * 0.05, + 0.1			

Table 4: Second hypothesis, progressive probabilities – regression summary 1

	Estimate	Std. error	t value	P(>|t|)	
δ_0	0.243	0.008	29.173	< 2.00*e^{-16}	***
δ_1	0.049	0.015	3.212	4.83*e^{-03}	**
R^2	0.364	Significance codes: *** 0.001, ** 0.01, * 0.05, + 0.1			

Table 5: Second hypothesis, progressive probabilities – regression summary 2

	Estimate	Std. error	t value	P(>|t|)	
δ_0	0.235	0.005	47.439	< 2.00*e^{-16}	***
δ_1	0.075	0.009	7.926	2.80*e^{-07}	***
R^2	0.777	Significance codes: *** 0.001, ** 0.01, * 0.05, + 0.1			

Table 6: Second hypothesis, progressive probabilities – different values of k

k	$\widehat{\delta_1}$	Std. error	P(>|t|)		R^2
5	0.049	0.015	4.83*e^{-03}	**	0. 364
6	0.075	0.009	2.80*e^{-07}	***	0. 777
7	0.103	0.017	1.33*e^{-05}	***	0.661
8	0.102	0.021	1.11*e^{-04}	***	0.572
9	0.104	0.024	3.47*e^{-04}	***	0.518
10	0.108	0.020	3.06*e^{-05}	***	0.629
15	0.170	0.029	1.64*e^{-05}	***	0.653
20	0.245	0.037	6.67*e^{-06}	***	0.711
		Significance codes: *** 0.001, ** 0.01, * 0.05, + 0.1			

Table 7: Second hypothesis, lower importance of earlier periods – regression summary

	Estimate	Std. error	t value	P(>|t|)	
δ_0	0.244	0.007	34.285	< 2.00*e^{-16}	***
δ_1	0.059	0.012	4.949	1.04*e^{-04}	***
R^2	0.576	Significance codes: *** 0.001, ** 0.01, * 0.05, + 0.1			

Table 8: Second hypothesis, lower importance of earlier periods – different values of k

k	$\widehat{\delta_1}$	Std. error	P(>|t|)		R^2
5	0.059	0.012	$1.04*e^{-04}$	***	0. 576
6	0.063	0.011	$4.65*e^{-05}$	***	0.611
7	0.054	0.007	$1.09*e^{-06}$	***	0.742
8	0.072	0.011	$2.64*e^{-06}$	***	0.715
9	0.087	0.016	$4.63*e^{-05}$	***	0.612
10	0.077	0.014	$2.63*e^{-05}$	***	0.635
15	0.105	0.019	$3.97*e^{-05}$	***	0.618
20	0.132	0.023	$2.18*e^{-05}$	***	0.642
		Significance codes: *** 0.001, ** 0.01, * 0.05, + 0.1			

Table 9: Second hypothesis – robustness test with different core process

k	$\widehat{\delta_1}$	Std. error	P(>|t|)		R^2
5	0.060	0.013	$1.31*e^{-03}$	**	0.578
6	0.067	0.021	$4.95*e^{-07}$	***	0.601
7	0.065	0.013	$5.84*e^{-12}$	***	0.701
8	0.048	0.010	$2.89*e^{-15}$	***	0.725
9	0.130	0.018	$9.75*e^{-07}$	***	0.608
10	0.115	0.016	$1.52*e^{-08}$	***	0.659
15	0.140	0.025	$4.88*e^{-13}$	***	0.588
20	0.153	0.008	$4.84*e^{-07}$	***	0.596
		Significance codes: *** 0.001. ** 0.01. * 0.05. + 0.1			

Table 10: Third hypothesis, same-sign shifts – regression summary

	Estimate	Std. error	t value	P(>|t|)	
δ_0	0.151	0.013	11.93	$5.54*e^{-10}$	***
δ_1	0.237	0.022	10.74	$2.95*e^{-09}$	***
R^2	0.865	Significance codes: *** 0.001, ** 0.01, * 0.05, + 0.1			

Table 11: Third hypothesis, same-sign shifts – different values of k

k	$\widehat{\delta_1}$	Std. error	P(>|t|)		R^2
5	0.237	0.022	$2.95*e^{-09}$	***	0. 865
6	0.255	0.284	$4.47*e^{-08}$	***	0.818
7	0.306	0.024	$2.42*e^{-10}$	***	0.898
8	0.339	0.034	$7.59*e^{-09}$	***	0.850
9	0.309	0.031	$9.25*e^{-09}$	***	0.847
10	0.407	0.046	$6.40*e^{-08}$	***	0.811
15	0.554	0.045	$3.15*e^{-10}$	***	0.895
20	0.646	0.064	$8.35*e^{-09}$	***	0.849

Significance codes: *** 0.001, ** 0.01, * 0.05, + 0.1

Table 12: Third hypothesis, sign-varying shifts – regression summary

	Estimate	Std. error	t value	P(>|t|)	
δ_0	0.194	0.004	50.57	$< 2.00*e^{-16}$	***
δ_1	-0.007	0.007	-1.12	0.279	
R^2	0.065	Significance codes: *** 0.001, ** 0.01, * 0.05, + 0.1			

Table 13: Third hypothesis, sign-varying shifts – different values of k

k	$\widehat{\delta_1}$	Std. error	P(>|t|)		R^2
5	-0.007	0.007	0.279		0. 065
6	-0.017	0.010	0.097	+	0.146
7	0.009	0.011	0.441		0.033
8	-0.019	0.017	0.279		0.065
9	-0.014	0.009	0.144		0.115
10	0.004	0.015	0.782		0.004
15	-0.008	0.023	0.740		0.006
20	-0.004	0.024	0.883		0.001

Significance codes: *** 0.001, ** 0.01, * 0.05, + 0.1

Table 14 Third hypothesis, manipulating the size of shifts – regression summary

	Estimate	Std. error	t value	P(>|t|)	
δ_0	0.177	0.007	27.14	$4.69{*}e^{-16}$	***
δ_1	0.067	0.011	5.89	$1.42{*}e^{-05}$	***
R^2	0.658	Significance codes: *** 0.001, ** 0.01, * 0.05, + 0.1			

Table 15 Third hypothesis, manipulating the size of shifts – different values of k

k	$\widehat{\delta_1}$	Std. error	P(>|t|)		R^2
5	0.067	0.011	$1.42{*}e^{-05}$	***	0.658
6	0.060	0.013	$1.77{*}e^{-04}$	***	0.551
7	0.067	0.011	$9.11{*}e^{-06}$	***	0.674
8	0.071	0.015	$1.87{*}e^{-04}$	***	0.549
9	0.118	0.020	$1.06{*}e^{-05}$	***	0.669
10	0.065	0.019	$3.29{*}e^{-03}$	**	0.389
15	0.137	0.024	$2.15{*}e^{-05}$	***	0.643
20	0.146	0.396	$1.68{*}e^{-03}$	**	0.431
		Significance codes: *** 0.001, ** 0.01, * 0.05, + 0.1			

Table 16 Third hypothesis – robustness test with different core process

k	$\widehat{\delta_1}$	Std. error	P(>|t|)		R^2
5	0.060	0.013	$9.59{*}e^{-07}$	***	0.561
6	0.067	0.021	$8.80{*}e^{-04}$	***	0.571
7	0.065	0.013	$4.23{*}e^{-07}$	***	0.632
8	0.048	0.010	$1.96{*}e^{-06}$	***	0.545
9	0.130	0.018	$9.10{*}e^{-14}$	***	0.597
10	0.115	0.016	$5.20{*}e^{-13}$	***	0.611
15	0.140	0.025	$5.98{*}e^{-09}$	***	0.583
20	0.153	0.008	$< 2.00{*}e^{-16}$	***	0.594
		Significance codes: *** 0.001, ** 0.01, * 0.05, + 0.1			

Table 17: Comparison of results for different total probabilities of breaks' occurence

p_t =	1%		2%		3%		4%	
k	$\widehat{\delta_1}$	ΔDoC	$\widehat{\delta_1}$	ΔDoC	$\widehat{\delta_1}$	ΔDoC	$\widehat{\delta_1}$	ΔDoC
5	0.062	0.322	0.075	0.318	0.069	0.286	0.082	0.291
6	0.063	0.345	0.052	0.351	0.063	0.315	0.059	0.298
7	0.072	0.372	0.083	0.356	0.068	0.309	0.081	0.301
8	0.077	0.392	0.084	0.387	0.072	0.343	0.079	0.331
9	0.097	0.377	0.094	0.405	0.101	0.352	0.085	0.324
10	0.082	0.393	0.092	0.385	0.081	0.358	0.118	0.373
15	0.123	0.422	0.135	0.410	0.139	0.411	0.142	0.362
20	0.132	0.445	0.145	0.439	0.157	0.425	0.151	0.384

Table 18: Comparison of results for different ratios between variances

σ_S^2/σ_c^2 =	0.2		0.3		0.4		0.5	
k	$\widehat{\delta_1}$	ΔDoC	$\widehat{\delta_1}$	ΔDoC	$\widehat{\delta_1}$	ΔDoC	$\widehat{\delta_1}$	ΔDoC
5	0.062	0.322	0.074	0.273	0.092	0.340	0.098	0.294
6	0.063	0.345	0.077	0.309	0.075	0.308	0.078	0.354
7	0.072	0.372	0.089	0.405	0.100	0.382	0.116	0.405
8	0.077	0.392	0.104	0.348	0.108	0.409	0.112	0.416
9	0.097	0.377	0.113	0.358	0.125	0.423	0.137	0.338
10	0.082	0.393	0.098	0.357	0.117	0.400	0.127	0.382
15	0.123	0.422	0.150	0.407	0.178	0.389	0.202	0.396
20	0.132	0.445	0.161	0.457	0.197	0.440	0.227	0.454