



Order patterns, their variation and change points in financial time series and Brownian motion

Christoph Bandt¹

Received: 21 October 2019 / Revised: 11 March 2020 / Published online: 20 March 2020
© The Author(s) 2020

Abstract

Order patterns and permutation entropy have become useful tools for studying biomedical, geophysical or climate time series. Here we study day-to-day market data, and Brownian motion which is a good model for their order patterns. A crucial point is that for small lags (1 up to 6 days), pattern frequencies in financial data remain essentially constant. The two most important order parameters of a time series are turning rate and up-down balance. For change points in EEG brain data, turning rate is excellent while for financial data, up-down balance seems the best. The fit of Brownian motion with respect to these parameters is tested, providing a new version of a forgotten test by Bienaymé.

Keywords Order pattern · Time series · Permutation entropy · Stock data

Mathematics Subject Classification 62M10 · 91B84 · 60G18

1 Overview

Given a finite series of data, time series analysis attempts to find laws of the mechanism or process which generated the data. One tool to this end are frequencies of order patterns, defined in Sect. 3. For a first impression, Fig. 1 shows the six order patterns of length 3. Given a time series x_1, \dots, x_T , a point t represents the pattern $\pi = 123$ if $x_t < x_{t+1} < x_{t+2}$. The frequency p_π of a pattern π is the number of time points which represent π , divided by $T - 2$. Such frequencies are estimates of probabilities of an underlying random process. They can be combined in various ways. The permutation entropy $-\sum_{\pi} p_{\pi} \log p_{\pi}$ measures the variety of patterns (Bandt and Pompe 2001). Applications to brain and heart data concern epilepsy (Bruzzo et al. 2008; Ferlazzo 2014), Alzheimer's disease (Morabito et al. 2012), effect of anaesthesia (Olofsen et al.

✉ Christoph Bandt
bandt@uni-greifswald.de

¹ Institute of Mathematics, University of Greifswald, 17487 Greifswald, Germany

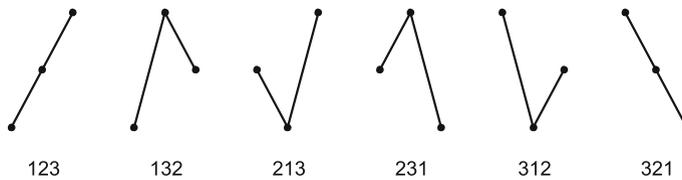


Fig. 1 The six order patterns of length 3

2008), and cardiac dynamics (Parlitz et al. 2012; Chicote et al. 2016; McCullough et al. 2017). For an overview of applications to physics, geophysics, environmental and climate data see Amigo et al. (2013), Amigo et al. (2015), and Zanin et al. (2012).

Moreover, probabilities of specific order patterns can give important information. For chaotic dynamical systems, certain patterns will not appear (Amigo 2010). Different versions of permutation entropy and the number of forbidden patterns were used to quantify and monitor the efficiency of stock (Zunino et al. 2009) and bond markets (Zunino et al. 2012, 2016). The probabilities of patterns can be combined to form correlation functions (Bandt 2015, 2017). The dependence of economical time series was studied in Schnurr (2014) by means of order patterns. Statistical theory of order pattern estimators was developed recently by Schnurr and Dehling (2017) and Betken et al. (2019), earlier work by Bandt and Shiha (2007) and Sinn and Keller (2011) was restricted to Gaussian processes.

Several authors have studied change points and segmentation of time series by means of order patterns (Sinn et al. 2013; Unakafov and Keller 2018). The interesting approach in Schnurr (2014) and Schnurr and Dehling (2017) uses bivariate data while we shall study the univariate case only. A very impressive example is sleep stage classification using high-frequency data of a single EEG channel (Nicolaou and Georgiou 2012; Kuo and Liang 2011; Bandt 2017, 2019). Here we have an abundance of data and few statistical problems. In the next section we will briefly describe this application.

Our main theme are order patterns in day-to-day financial data and their statistics. The zigzag patterns of up and down have long been considered by stock traders. Today they are certainly incorporated in secret algorithms for high-frequency trade. We are interested in the statistical error of estimates of pattern probabilities, and in the change point problem. We introduce our standard example of oil prices in Sect. 4.

A curious feature of such examples is their apparent ‘self-similarity’ studied by Mandelbrot since the 1960s (Mandelbrot 1997). This property implies equality of frequencies of patterns for different lags, which can be compared with simulated samples of Brownian motion and other Levy processes. One goal of the paper is to develop tests which decide whether Brownian motion is an appropriate model. In Sect. 5 this is done by simulation, in Sect. 7 rigorously.

It seems that for a large part, ‘order self-similarity’ is due to the uneven sampling of market data, with missing weekends and holidays. In contrast to measurements done in nature, the observed values are not varying in natural time. Their change is triggered by trade. Volume and number of buying and selling orders provide alternative scales to natural time. Under such conditions, classical tools like autocorrelation become

useless. It may be a necessity to postulate equality of pattern frequencies for small lags. Our discussion in Sect. 6 shows that the assumption is justified and can be used to improve estimates of pattern frequencies.

Our expectations should be modest: with at most a few thousand data, statistical accuracy will not be magic. In Sect. 7 we introduce the two most stable and interpretable order parameters, up-down balance and turning rate. They are used to provide statistical tests for comparing models and data series. Actually, a first test for the turning rate was suggested already in Bienaymé (1874, 1875). The basic message is that pattern statistics has the $1/\sqrt{n}$ accuracy coming from binomial distribution. In Bienaymé's case it is even better because of negative correlations. In Sect. 9 we use order patterns to determine change points in financial time series. It turns out that up-down balance is the most appropriate parameter.

2 A big data series from medicine

We start with a brief description of segmentation of big series of brain data by means of the order structure. This has been done by many authors, cf. Olofsen et al. (2008), Ouyang et al. (2010), Nicolaou and Georgiou (2012), Kuo and Liang (2011), Sinn et al. (2013), Bandt (2017), and Unakafov and Keller (2018). Our version in Bandt (2018) is particularly simple. We considered electroencephalographic (EEG) sleep data measured by Terzano et al. (2001), and publicly available at *pysionet* (Goldberger et al. 2000). Healthy volunteers and patients were measured for one night in hospital with a sample rate of 500 Hz. These are multivariate data, as is the standard in sleep labs, and sleep stages (wake—REM sleep—sleep stages 1 to 4) were annotated by trained experts for every epoch of 30 s. For our classification, we used only two EEG channels, each with 20 million data points for one night.

We determined a score for each epoch in an extremely simple way: we just counted the relative number of local maxima and minima in the time series. In Fig. 2 it can be seen that the result almost coincides with the expert's annotation. In white noise, the turning rate—the relative amount of local extrema—is known to be $\frac{2}{3}$ (Bienaymé 1874). In Fig. 2 we see turning rates between 0.35 and 0.45 when the proband is awake, with larger rates in the more frontal channel. In REM sleep, rates were between 0.3 and 0.35, also larger for the more frontal channel. When the patient was in non-REM sleep, turning rates went below 0.3, with larger values in the central channel. The smallest turning rate is below 0.2, in the first phase of deep sleep. After midnight, turning rates of sleep stages 3 and 4 gradually increased. Such results, with individual differences, were observed for about 20 persons measured by Terzano et al. (2001) with 500 Hz while for lower sampling rate, and for other data contaminated by 50 Hz noise from the power net, the results were less impressive (Bandt 2018).

Here we looked for many change points, and a very simple method was sufficient. The main thing was to express the structure by the appropriate parameter (see Sect. 7 for details). Of course, this is preliminary research, and much work has to be done to make it an applicable method. But we are in a comfortable position with 15000 data for each epoch. For this application, a precision of ± 10 s for change points is fine, so an error of ± 1000 time points does not count. The data contain plenty of artefacts

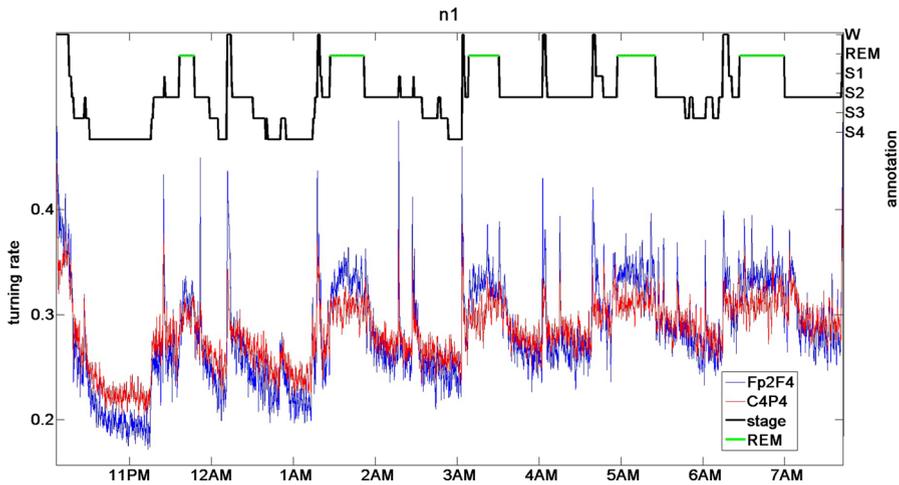


Fig. 2 Sleep stage classification by turning rate. The step function at the top represents an expert’s annotation, given for epochs of 30 s. At the bottom, the average turning rate is drawn for every epoch, for two EEG channels. The surprising coincidence was found for the other patients, too

when the patient moves, rolls with the eyes in dream, gnabs with the teeth etc. A few bad epochs can be tolerated, however. Most of them are annotated ‘wake’ in Fig. 2. The sliding window technique works well. We just determine turning rate for epochs of 30 s. For greater detail, we can take overlapping windows, with a shift of one second only (Bandt 2018). There are no serious statistical problems.

In this paper on historic economical data, the situation is quite different. We have a few thousand values altogether. We must think about which structural parameters can be estimated with reasonable accuracy, and we have to deal with statistical errors.

3 Pattern frequencies as one facet of a random process

As we said, we try to find properties of the process which generates the data. We begin with a brief glance at the theory. A stationary random process X_1, X_2, \dots in discrete time, $\mathbb{N} = \{1, 2, \dots\}$ is essentially defined by its distribution, a probability measure P on Borel sets B of the space $\Omega = \mathbb{R}^{\mathbb{N}}$ of all possible infinite time series y_1, y_2, \dots , which is invariant under time shifts. It is enough to know the probabilities of those Borel sets $B \subset \mathbb{R}^n$ which depend only on the first values y_1, \dots, y_n , for arbitrary n .

A statistician will ask how probabilities $P(B)$ are estimated from a finite time series $x = x_1, x_2, \dots, x_T$. Since we assume stationarity, an estimate $\widehat{P}(B)$ is the relative frequency of those vectors $(x_{t+1}, \dots, x_{t+n})$ which belong to the set B , for $t = 0, \dots, T - n$. As a formula,

$$\widehat{P}(B) = \frac{1}{T + 1 - n} \#\{t \mid 0 \leq t \leq T - n, (x_{t+1}, \dots, x_{t+n}) \in B\}. \tag{1}$$

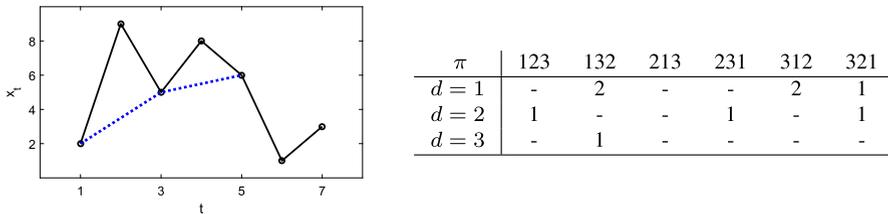


Fig. 3 Example time series and order pattern frequencies. The dotted line indicates $d = 2$.

Here $\#A$ denotes the cardinality of a set A . Let us assume that the stationary process is ergodic. Then a classical theorem of Birkhoff says that for $T \rightarrow \infty$, the estimate $\widehat{P}(B)$ will converge to the true value $P(B)$, with probability one.

Alright! But in practice, T is fixed and fairly small, say $T = 1000$. To get reasonable estimates, one usually takes sets B like $\{x|x_1 < b\}$ for given numbers b , which leads to the one-dimensional distribution. Instead of considering multidimensional Borel sets, it is preferred to study autocorrelation, based on averages over t of functions $x_{t+1}x_{t+n}$ for $n = 1, 2, \dots$

There is, however, one class of multidimensional sets B for which estimation makes sense: the sets $\{x|x_j < x_k\}$ and their intersections. The set $\{x|x_1 < x_3\}$ describes a halfplane in the x_1x_3 -plane, and $B = \{x|x_1 < x_2 < x_3\}$ describes a ‘cone’, both in \mathbb{R}^3 and Ω . The latter set B is symbolized as the first order pattern in Fig. 1. There are five other patterns or permutations π . For example, $\pi = 231$ corresponds to $B = \{x|x_3 < x_1 < x_2\}$.

Note that a permutation π of length n is a one-to-one mapping from $\{1, \dots, n\}$ onto itself. Figure 1 shows the graphs of the six permutations of length 3. We shall say that n equally spaced values $x_{t+d}, x_{t+2d}, \dots, x_{t+nd}$ in our time series represent the permutation π of length n if they are in the same order. That is, $\pi(j) < \pi(k)$ if and only if $x_{t+jd} < x_{t+kd}$, for $1 \leq j < k \leq n$. Now we can count relative frequencies of permutations in the time series, which we shall denote $p_\pi(d)$. Let

$$p_\pi(1) = \frac{1}{T+1-n} \#\{t \mid 0 \leq t \leq T - n, (x_{t+1}, x_{t+2}, \dots, x_{t+n}) \text{ realizes } \pi\}. \quad (2)$$

The case $d = 1$ corresponds to successive values in the series, but sometimes it is good to compare values which are 2 or d steps apart. The Formula (1) for sets B determined by a permutation π of length n and a lag $d \geq 1$ reads as follows.

$$p_\pi(d) = \frac{1}{T+d-nd} \#\{t \mid -d + 1 \leq t \leq T - nd, (x_{t+d}, x_{t+2d}, \dots, x_{t+nd}) \text{ realizes } \pi\}. \quad (2d)$$

Only formally, we allow t to be negative. Otherwise we had to consider permutations on $0, 1, \dots, n - 1$ as in Amigo et al. (2015). Figure 3, taken from Bandt (2017), shows a small example which leads to the frequencies $p_{132}(1) = \frac{2}{5}$, $p_{321}(2) = \frac{1}{3}$, and $p_{321}(3) = 0$.

We have to note that our method excludes equality of values $x_j = x_k$. In theory, equality happens only with probability zero when the one-dimensional distribution of the underlying process has a density function. In practice, equality of values will

happen, but will be rare when the values have an accuracy of 5 or more decimals. As explained in Bandt (2019), pairs of equal values, as well as missing values, can just be disregarded in the calculation. A simpler practical method to eliminate equal values is to add a tiny white noise (uniform random numbers times 10^{-7} , say) to the series x .

There are at least three good arguments for considering frequencies of such order patterns. First, since they describe cones in Ω , their estimates are robust with respect to noise and outliers, and not changed by an increasing nonlinear transformation of the data. Next, order patterns can be evaluated very easily and extremely fast by computer. There are few possible sources of error.

The most important argument for the use of order patterns are the weak stationarity assumptions which are required. The process X_1, X_2, \dots need not be stationary. It need only have stationary increments in the sense that for all $t, n \geq 1$

$$(X_{t+1} - X_t, \dots, X_{t+n} - X_t) \text{ has the same distribution as } (X_1, \dots, X_n). \quad (3)$$

For pairwise different numbers x_t, \dots, x_{t+n} , the vector $(x_{t+1} - x_t, \dots, x_{t+n} - x_t)$ always represents the same permutation as the vector $(x_{t+1}, \dots, x_{t+n})$. The same argument works for arbitrary d . We obtain the following conclusion.

Proposition 1 *For an ergodic process X_1, X_2, \dots with stationary increments, $p_\pi(d)$ in (2) and (2d) is a consistent and unbiased estimator of the probability that the vector $(X_{t+d}, \dots, X_{t+nd})$ represent the permutation π , for any fixed t .*

This is a special case of the estimator (1) which is known to be unbiased. The proposition says that our naive formula (2d) makes sense. In the sequel, we consider only patterns of length $n \leq 4$.

Financial data have a random walk structure and are typical examples of processes with stationary increments. Brownian motion is their classical model. It is not stationary, and its autocorrelation function depends on two arguments. Its pattern frequencies as functions of the lag d can be analytically determined, however Bandt and Shiha (2007).

4 Patterns of length four in a series of oil prices

Permutations of length 2 and 3 lead to two important parameters β and τ , as discussed in Sect. 7 below. To differentiate between various time series, we can use all 24 permutations of length 4. They provide detail information, and can be reasonably estimated when more than 1000 values are available. For a study of the 120 permutations of length 5 we would need much longer series.

For convenience, we shall assign to each permutation a number between 1 and 24, according to lexicographic order, as indicated in Table 1.

Figure 4 presents order four frequencies (our abbreviation of ‘frequencies of order patterns of length 4’) for daily closing oil prices of the brand West Texas Intermediate (WTI), a standard reference for American crude oil. The series, given in Fig. 5 below, contains 8497 values ranging from January 1986 up to September 2019. The $p_\pi(d)$

Table 1 Numbering of permutations of length 4

1	2	3	4	5	6	7	8	9	10	11	12
1234	1243	1324	1342	1423	1432	2134	2143	2314	2341	2413	2431
13	14	15	16	17	18	19	20	21	22	23	24
3124	3142	3214	3241	3412	3421	4123	4132	4213	4231	4312	4321

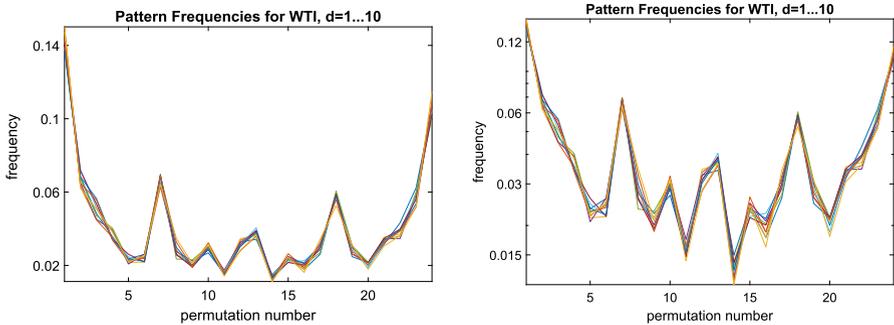


Fig. 4 Order four frequencies for the oil price series WTI, 1986–2019. Order patterns are numbered from 1 to 24. Their frequencies are shown on linear scale at the left, and on logarithmic scale at the right. Different curves correspond to the lags $d = 1, \dots, 10$. Apparently, the frequencies do not depend on the lag

are shown for the lags $d = 1, \dots, 10$. The most apparent fact is that frequencies for different lags are very similar at each pattern. This will be discussed in the sequel.

On the left, frequencies are shown on linear scale. The largest frequencies appear for patterns 1234 and 4321, with $p_\pi(d)$ around $1/8$. This happens for all daily price data. Here, p_{1234} is clearly larger $1/8$ and p_{4321} is smaller, which can be explained by the increasing trend of the series. The largely increasing patterns 1243 and 2341 and their decreasing counterparts 4312, 3214 have probabilities near to $1/16 = 0.0625$. All other patterns have smaller probabilities.

On the right, pattern frequencies are shown on logarithmic scale, in order to distinguish better the rare patterns. Differences here actually represent quotients, and deviations from the mean have to be interpreted as relative errors. The pattern with the smallest frequency, less than 1.5 percent, is 3142, followed by 2413 with almost 2 percent. Note that the average frequency is $1/24 \approx 4.2\%$. The patterns in the middle, with numbers 8 to 17, as well as 5,6 and 19,20, are below average.

The figure shows an obvious symmetry between patterns with number k and $25 - k$. Actually, pattern number $25 - k$ is obtained by reversing all order relations in pattern number k . Rank 4 is interchanged with 1, rank 2 with 3, and $<$ with $>$. The graph of pattern $25 - k$, drawn as in Fig. 1, is the graph of the pattern with number k turned upside down. It seems plausible that these two patterns have the same frequency, but in our series it is not quite true. The permutations with higher number have somewhat smaller frequency than their counterparts, with exception of number 12 and 13: 3124 is more frequent than 2431. Can this break in symmetry be explained by a trend?

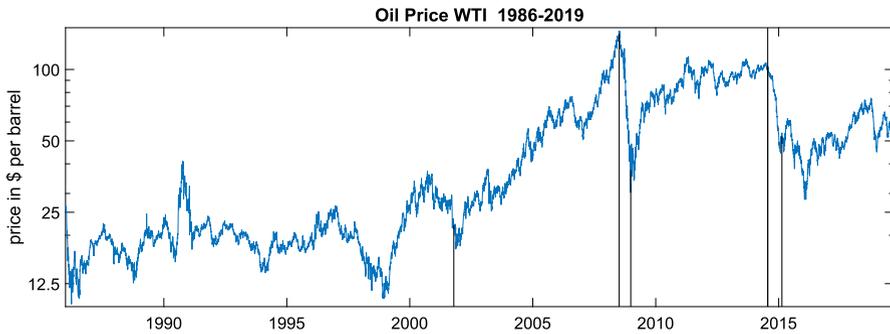


Fig. 5 The time series of WTI oil prices on logarithmic scale. To find out how pattern frequencies are determined by a trend, an ad hoc segmentation is defined. It contains two periods of increasing prices and two intervals without clear trend. Periods of decrease were too short to provide reliable pattern estimates

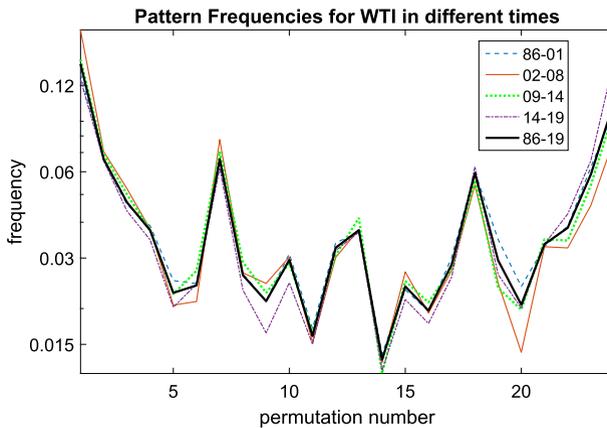


Fig. 6 The pattern frequencies of WTI in different time periods. Influence of trend is small

The time series of $T = 8497$ oil prices, drawn in Fig. 5 on logarithmic scale, shows a general upwards trend, as most price series do. It is interrupted by two rather sudden drops of prices: one in 2008 when the world economy was hit by the financial crisis, and one in 2014 when some producers, notably Saudi Arabia, increased their supply to compete with the upcoming fracking industry in the USA.

So the series is not stationary and needs segmentation, which will be discussed in Sects. 9 and 10. To study the effect of the trend on pattern frequencies, we define a preliminary segmentation into four pieces by eyesight. The first subseries with 4000 values covers the time from 2 January 1986 to 16 October 2001 where no trend can be seen (the first price is 25.56\$, the last one only 22.01 \$). The second series with 1680 values describes the time of strong price increase from 17 October 2001 up to 7 July 2008 with the maximum value. Another series with increasing trend and 1400 values covers the time from 26 December 2008 to 22 July 2014. The last series consists of the remaining 1147 values from 23 July 2014 to 3 September 2019, with almost no trend. The times of rapidly decreasing prices, with 120 business days in 2008 and 150 in 2014, were too short to provide a reliable statistics of pattern frequencies.

Order four pattern frequencies for the four subseries are shown in Fig. 6. They do not look as different as expected. It is true that for the strongly increasing series 2009–2014, the increasing patterns on the left side of the figure have larger frequencies and the decreasing patterns on the right side, with number 20 to 24, say, have smaller frequencies. And for the time 2014–2019 without trend, pattern 4321 appears more often than 1234. But some of the frequencies, notably number 11 to 14 in the middle, are almost constant through time. On the whole, frequencies do not differ much.

How can we decide about differences of such frequencies with statistical rigor? This will be a theme of the next sections. First we introduce a standard model.

5 Brownian motion as model for patterns in financial data

Since the work of Bachelier in 1900, Brownian motion (abbreviated BM) is a basic model for financial time series. It is usually taken to describe the logarithm of prices. For day-to-day data it postulates that the difference $R_t = \log x_{t+1} - \log x_t$, the so-called log return of day t , has a Gaussian distribution with mean zero and some variance σ^2 . Moreover, log returns of different days should be independent.

The independence assumption is plausible. When there would be a specific dependence between returns of today and tomorrow, some speculators would immediately bet against it to earn money, which would cause the dependence to disappear. The Gaussian distribution is not realistic, however, since big returns, both negative and positive, are more frequent in practice than the normal distribution allows (Mandelbrot 1997). Moreover, the important phenomenon of varying volatility is not contained in the Brownian motion model.

For order patterns, Brownian motion is a good model. It makes no difference whether we model logarithms or original values, because $\log x$ is an increasing function. The value of σ is also irrelevant, so we take $\sigma = 1$. Thus we let x_t , $t = 1, \dots, T$ be a realization of standard Brownian motion—the cumulative sum of a series of standard normal random numbers. It can be simulated by the Matlab command `x=cumsum(randn(1,T))`.

Pattern frequencies for lags $d = 1, \dots, 10$ of a typical simulation of Brownian motion with length $T = 8500$ are shown in Fig. 7. Theoretically, patterns k and $25 - k$ appear with the same probability, but this need not be true in a simulation. Moreover, the process of Brownian motion in discrete time is known to be self-similar. More precisely, B_1, B_2, B_3, \dots and $(B_s, B_{s+d}, B_{s+2d}, \dots)/\sqrt{d}$ have the same distribution for any lag $d \geq 1$ and any initial time s . As a consequence, the probabilities $p_\pi(d)$ do not depend on the lag d . Of course, this need not be so in a simulation. For large length $T = 10^6$ we checked that the coincidence is convincing for virtually every simulation. Here we want to see what happens for $T = 8500$, the same length as our oil price data. Comparing Figs. 7 and 4, we see that the coincidence of frequencies for different lags in the data is not worse than in the model, where it should be perfect by definition. We try to clarify this point in the next section.

For Brownian motion, probabilities of order 4 patterns can be determined analytically (Bandt and Shiha 2007): since the multivariate standard normal distribution is spherically symmetric, octant probabilities can be determined as surfaces of spherical

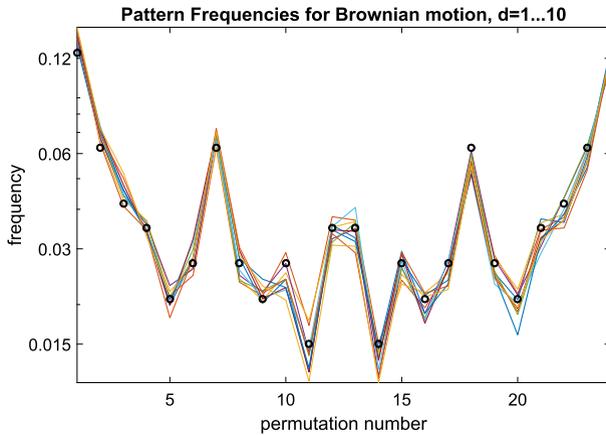


Fig. 7 A simulation of Brownian motion pattern frequencies for lags $d = 1, \dots, 10$. We took $T = 8500$ as for the WTI series above. Theoretical pattern probabilities are marked by circles

Table 2 Exact probabilities of permutations of length 4 for Brownian motion

Pattern number	1	2,7	3	4,12	5,9	6,8,10	11
Probability	1/8	1/16	1/24	0.035	1/48	0.027	0.015

Patterns k and $25 - k$ have the same probability

Numbers with decimal point indicate irrationals involving $\arcsin 1/\sqrt{3}$

triangles. The angles of these triangles are related to certain correlation coefficients. Table 2 lists the results of Bandt and Shiha (2007). It is curious that there are seven different probabilities of which only four are rational.

Comparing the distributions of pattern frequencies of WTI prices in Fig. 4 and BM in Fig. 7, we see little difference. Is BM a good model in our case? We consider distributions of order 4 patterns as vectors in \mathbb{R}^{24} and study their Euclidean distances.

Let $q = (q_1, \dots, q_{24})$ denote an empirical distribution of WTI for one of the four time periods in Fig. 6. We determine the distance $\|q - \bar{q}\|$ to the distribution \bar{q} of the WTI over the whole time, and the distance $\|q - b\|$ to the theoretical distribution b of Brownian motion given in Table 2. Moreover, $N = 10^5$ simulations of BM with the sample size of the respective WTI series for q are performed and their pattern distributions b_k calculated. For $k = 1, \dots, N$ we determine $d_k = \|b_k - b\|$.

The distribution of all d_k serves as test distribution for the null hypothesis that q is obtained from the Brownian motion model. It resembles a chi-distribution for 23 degrees of freedom, with more heavy tail on the right, but not far from normal. Our simulation provides the median of the d_k , and an estimate of the p -value of the data:

$$p = \frac{1}{N} \#\{k \mid 1 \leq k \leq N, d_k > \|q - b\|\}.$$

For different lags d , the distances $\|q - b\|$ can differ by 20 or 50 %, but they were always within a factor of 2. The resulting p -value, however, can differ by a factor

Table 3 Distances of pattern distribution q of WTI prices in different periods to their mean \bar{q} and to the exact distribution b of BM

WTI time series	1986–2019	1986–2001	2001–2008	2009–2014	2015–2019
Number of values	8497	4000	1680	1400	1150
Distance $\ q - \bar{q}\ $ to whole WTI	0	.013	.047	.014	.014
Distance $\ q - b\ $ to BM	.025	.024	.070	.034	.022
Median of distances $\ b_k - b\ $.011	.016	.024	.026	.029
p -value in %	0.04	3.8	<0.01	13.6	88.9

This was compared with the distance of 10^5 simulated sample distributions b_k of BM to b . During the last two periods, Brownian motion fits well

10 and more. Thus for a distribution with 24 frequencies, the dependence on the lag can be a problem. This experience leads us to simpler parameters of the time series, introduced in Sect. 7.

Here we decided to take the mean of the frequencies for lags 1, 2, and 3 as our empirical distribution q . This method is justified in the next section. The results are shown in Table 3. It turns out that for the last period 2015–2019, BM is a perfect model with $p = 88.9\%$. The patterns of the third period 2009–2014 could also well be generated from BM. For the whole series as well as for the first two periods, there are significant differences, however.

What is the reason for the differences? In the second period, it is the strongly increasing trend of the prices, which results in large frequencies p_{1234} and small frequencies p_{4321} . This is not surprising and could be mended by taking a biased BM with linear trend fitted to the increase of prices in that time. For the whole series, the patterns would also fit better to a BM with linear trend. However, in that case the trend has to be chosen in such a way that today's prices would range above 300 instead of 60 \$. Thus there may be another structural difference between WTI and BM for the whole series and for the first rather stationary period, although the p -value of 3.8 % in Table 3 is not convincing. We return to this question later.

6 Order-self-similarity of financial data

Before we go on, we make a few general remarks. Instead of studying the prices x_t , it is common to consider the increments $Z_t = x_{t+1} - x_t$ or log returns $R_t = \log x_{t+1} - \log x_t$ of every day. Their distribution is determined, and usually has heavier tails than normal distribution. Here we study order patterns in the original data, however. Returns come in automatically when we compare values: $x_{t+1} > x_t$ means that the increment and log return of day t are positive. Our order patterns can be considered as combinations of signs of returns for different days, and different time lags.

In contrast to time series which are measured with a fixed sampling rate, daily financial data come with an **irregular time scale**. At weekends as well as on holidays, the stock market is closed. We pretend that Monday is the day following Friday, although we know that much business is going on throughout the weekend. Moreover, there are days where no trade takes place, and other days are very busy. It was suggested

that the time scale should be modulated by the trading volume, but this is not done. As a consequence, studies of daily market data by classical methods of time series analysis, as autocorrelation and frequency spectrum, do not reveal clear structure.

On the positive side, it has often been observed that financial series have a nice symmetry property. They are statistically self-similar in the sense that a magnification of the function at one place looks very similar to the original function at another place (Mandelbrot 1997). A random process in continuous time, $Y_t, t > 0$ is said to be *self-similar* if there is an exponent $H > 0$ such that

$$Y_{rt} = r^H Y_t \quad \text{in distribution, for each positive number } r. \quad (4)$$

For details, see Embrechts and Maejima (2002) or Embrechts et al. (1997), section 8.9. Brownian motion with $H = \frac{1}{2}$ is the standard example. Certain Levy Processes and fractional Brownian motion are also self-similar, and have been used as models of financial data. Usually, equation (4) is only considered for the one-dimensional distribution of the Y_t , and there are methods to determine the Hurst exponent H from data, see Mandelbrot (1997) and Sinn and Keller (2011). The mathematical definition applies to the whole process, however. Thus for every $n, d \geq 1$

$$(X_d, \dots, X_{nd}) \text{ has the same distribution as } d^H \cdot (X_1, \dots, X_n). \quad (5)$$

This multidimensional setting is practically impossible to check with statistical methods. Using order patterns, we now define a simpler concept. Note that (X_1, \dots, X_n) and $c \cdot (X_1, \dots, X_n)$ represent the same order pattern for each positive constant c .

So let us say that a process Y_1, Y_2, \dots is *order self-similar* if for every $n, d \geq 1$

$$(X_d, \dots, X_{nd}) \text{ has the same pattern distribution as } (X_1, \dots, X_n). \quad (6)$$

Proposition 2 *Each self-similar process is order self-similar. If the process is also ergodic and has stationary increments, then $p_\pi = \frac{1}{m} \cdot \sum_{d=1}^m p_\pi(d)$ is a consistent and unbiased estimator of the pattern probability of π , for any fixed $m \geq 1$.*

The proof follows from Proposition 1 and the remark above. Thus we can use order self-similarity to improve our estimators, as already done in the previous section. Moreover, order self-similarity is a parameter-free concept—we do not need a Hurst exponent. And it seems rather easy to check this concept with sufficiently many data.

Problem Self-similar processes are not stationary but their standard examples have stationary increments. Are there stationary order self-similar processes?

Why should a daily financial series be self-similar? There are small and big orders on the market. Some shareholders keep their items for a long time, others will sell them immediately. Briefly, there are actions of all sizes which produce structure of all sizes in the time series. This may be true, but in our opinion the irregular time sampling of daily financial series contributes a lot to their self-similarity.

Our time series x_t mixes time differences over one day, like Monday to Tuesday, with three-day differences from Friday to Monday. When we consider lag 3, like Monday to

Thursday, then most differences cover five days, like Thursday to Tuesday. With such an irregular scale, our only chance seems to postulate that order pattern probabilities do not depend on the lag, at least for small lags. This postulate is supported by empirical data like Fig. 4.

Henceforth, we shall assume that financial data fulfil order self-similarity in the sense of (6), for $d \leq 6$. We use this assumption to improve estimates by applying Proposition 2 with $m = 3$ or $m = 6$. It makes no sense to go to much larger m . On one hand, the data will not fulfil (6) for large d , on the other hand the variance of estimates $p_\pi(d)$ increases with d , as we shall see below.

7 Turning rate and up-down balance

Pattern frequencies are not so easy to interpret. Certain combinations of pattern frequencies are more meaningful. Let us fix the lag $d = 1$ for simplicity. The *permutation entropy* of order n is a sum over all patterns of length n .

$$H = - \sum_{\pi} p_{\pi} \log p_{\pi} , \quad (7)$$

This was introduced as a measure of complexity of the data (Bandt and Pompe 2001). See Amigo (2010), Zanin et al. (2012), Amigo et al. (2013), and Amigo et al. (2015) for details, applications and related concepts. Here we shall focus on two simpler quantities. For patterns of length 2, we consider the

$$\text{up-down balance} \quad \beta = p_{12} - p_{21} = 2p_{12} - 1 . \quad (8)$$

This parameter measures asymmetry, non-Gaussianity or irreversibility of the process (Bandt and Shiha 2007; Bandt 2015, 2019). If there is an increasing or decreasing trend in the series, then β will be positive or negative, respectively. In the extreme case of an increasing or decreasing time series β assumes its maximum 1 and minimum -1 , respectively. However, β can also be positive in a stationary process. For instance, a stock price could increase always from Monday to Friday, and then fall to the level of last Monday over the weekend. In that case, $\beta = \frac{4}{5} - \frac{1}{5} = \frac{3}{5}$. Many time series in nature, for example the daily number of sunspots (Bandt and Shiha 2007), show similar behaviour. A random walk, that is, a process with independent stationary increments, will have no trend if the increment $Z = X_{t+1} - X_t$ has zero mean. The median need not be zero, and in that case $\beta = 2P(Z > 0) - 1$ will be positive or negative. There are also asymmetric Levy processes where the mean of Z does not exist. An autoregressive process with exponential noise is stationary and has nonzero β for small d . An example is discussed at the end of this paper. For financial data, β often differs from zero, but not too much, as Table 4 indicates for our oil data.

Our other important parameter is based on patterns of length 3. It is the

$$\text{turning rate} \quad \alpha = p_{132} + p_{231} + p_{213} + p_{312} = 1 - p_{123} - p_{321} , \quad (9)$$

the relative amount of local maxima and minima (turning points) in the series. This is an intensity parameter which varies between 0 and 1. The extreme case zero is

Table 4 Comparison of two estimators of turning rate and up-down balance for different segments of the oil price series

WTI time series	1986–2019	1986–2001	2001–2008	2009–2014	2015–2019
Number of values	8497	4000	1680	1400	1150
Turning rate α^1	0.510	0.506	0.528	0.509	0.507
Turning rate α^3	0.502	0.512	0.488	0.500	0.499
z -value of α^1	1.84	0.76	2.29	0.67	0.47
Up-down balance β^1	0.032	0.025	0.072	0.044	0.050
Up-down balance β^3	0.044	0.038	0.112	0.063	0.035
z -value of β^1	2.95	1.58	2.95	1.65	1.69

z -values were calculated from Proposition 3 below

approached by a smooth time series which has very few turning points. The other extreme is an alternating series, like $x_t = (-1)^t$, for which each point is a turning point. For a series of independent random numbers (white noise) we have $\alpha = \frac{2}{3}$, and for Brownian motion $\alpha = \frac{1}{2}$.

In previous work (Bandt and Shiha 2007; Bandt 2015, 2019), the author considered the persistence $\tau = p_{123} + p_{321} - \frac{1}{3} = \frac{2}{3} - \alpha$ instead of α as a function of d . This function shares many properties with autocorrelation. It is zero for white noise, for instance. However, turning rate was essentially introduced in Bienaymé (1874) and is a very intuitive concept while the word ‘persistence’ is used with various meanings even in mathematics. So τ is not used here.

H , β , and α can be considered for arbitrary lag $d \geq 1$. So they become functions of d which can be used as correlation functions. This is very useful in the case of big data series from weather or medicine where certain periodicities appear on various scales (Bandt 2019). The result shown in Fig. 2 was obtained with the turning rate for $d = 4$. For financial data, which are approximately order self-similar, we have essentially only one value α for small d . But Proposition 2 provides different ways of calculation, in particular

$$\alpha^1 = \alpha(1), \quad \alpha^3 = \frac{1}{3} \sum_{d=1}^3 \alpha(d), \quad \text{and} \quad \alpha^6 = \frac{1}{6} \sum_{d=1}^6 \alpha(d).$$

The same holds for β . In Table 4 we compare α^1, β^1 with α^3, β^3 . We see that the turning rate of our oil data is $\frac{1}{2}$ in all periods, with very small fluctuation. The differences of β from zero are a bit larger. The z -values are obtained from Proposition 3 below.

8 The variance of α and β

Defined in terms of the estimators p_π , the quantities H , α , and β are in fact themselves estimators of respective parameters of the underlying process. For instance, β estimates

$b = P(X_1 < X_2) - P(X_1 > X_2)$. For an ergodic process with stationary increments, Proposition 1 implies that α and β are consistent and unbiased estimators. The entropy H , given by a continuous but nonlinear function of the p_π , is consistent but not unbiased.

Now we determine the quadratic error of these estimators. We calculate the variance of the functions α and β , taken over all possible realizations of Brownian motion. The following fact holds for all processes with independent and symmetrically distributed increments.

Proposition 3 *For samples of length T from Brownian motion and lag $d = 1$, the number of turning points and the number of up-steps both follow the statistics of tossing a fair coin. The distribution of the number of turning points is binomial with $n = T - 2$ and $p = \frac{1}{2}$. The number of t between 1 and $T - 1$ with $x_t < x_{t+1}$ has binomial distribution with $n = T - 1$ and $p = \frac{1}{2}$. As a consequence, we have*

$$E(\alpha) = \frac{1}{2}, \quad \text{Var}(\alpha) = \frac{1}{4(T - 2)}, \quad E(\beta) = 0, \quad \text{Var}(\beta) = \frac{1}{T - 1}.$$

Proof The increments $Z_t = X_{t+1} - X_t$ of BM are independent and have positive sign with probability $\frac{1}{2}$. Thus counting the number n_{12} of up-steps t with $Z_t > 0$ is really the same as tossing $T - 1$ fair coins and counting heads. Then $p_{12} = n_{12}/(T - 1)$ has mean $\frac{1}{2}$ and variance $\frac{1}{4(T - 1)}$, and the assertion for $\beta = 2p_{12} - 1$ follows.

A point t between 2 and $T - 1$ is a turning point (local maximum or minimum) of the time series x_1, \dots, x_T if the increments $z_{t-1} = x_t - x_{t-1}$ and $z_t = x_{t+1} - x_t$ have the same sign. The events to be counted are now $A_t = \{Z_{t-1}Z_t > 0\}$ for $t = 2, \dots, T - 1$. Since the signs of the Z_t are independent and are $+$ with probability $\frac{1}{2}$, we have $P(A_t) = \frac{1}{2}$ for each t . When we show that the A_t are independent, we do again have a series of fair coin tosses. From the definition of A_t , it is clear that A_t is independent on all A_s with $s \leq t - 2$. It remains to show that A_{t-1} and A_t are independent:

$$P(A_{t-1} \cap A_t) = P(Z_{t-2}, Z_{t-1} \text{ and } Z_t \text{ have the same sign}) = \frac{1}{4} = P(A_{t-1})P(A_t).$$

The rest of the proof is similar as for β . □

We apply the result to α^1 and β^1 in Table 4, to test the fit of Brownian motion to our oil price data. The z -value of α^1 is obtained as

$$z = (\alpha - \alpha_{BM})/\sigma(\alpha_{BM}) = (\alpha - \frac{1}{2}) \cdot 2\sqrt{T - 2},$$

for β^1 we calculate $z = \beta\sqrt{T - 1}$. We compare with the standard normal distribution. For turning rates, only the period 2001–2008 showed significant differences. The z -values of β are larger, but the significance for 2001–2008 and for the whole time series can be attributed to trend. On the whole, BM is a good model for order patterns of the oil prices.

Proposition 3 is paradigmatic because it shows that for all estimators $p_\pi(d)$ we can expect standard deviation and confidence bound of the form C/\sqrt{T} where C is

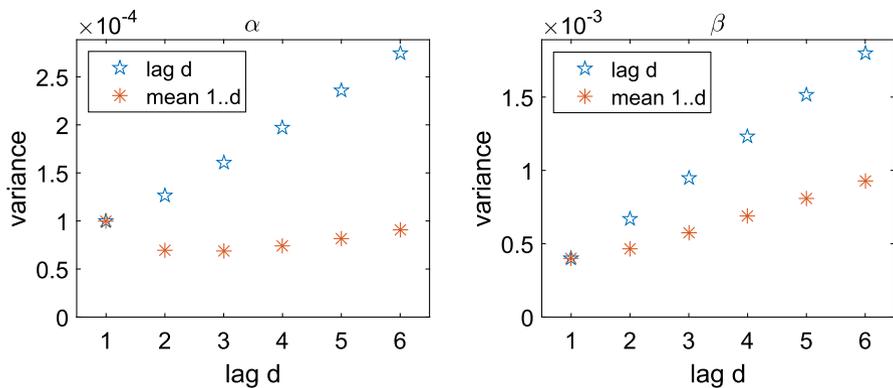


Fig. 8 Dependence of the variance of α and β on the lag d for BM, obtained by simulating 10^5 trajectories of BM with length $T = 2500$. The variance of α increases slowly, and an average over lags 1 to 3 diminish the variance of the estimator. For β the variance increases even in the average

a constant. However, beware of correlations! The proposition is not true for lags d greater than 1. The reason is that $Z_t(d) = X_{t+d} - X_t$ and $Z_{t+1}(d) = X_{t+1+d} - X_{t+1}$ are not independent since contain the common part $X_{t+d} - X_{t+1}$, and this term is likely to become the main term when d is large. So there is positive correlation among the sets A_t , which increases the variance. For BM this can be studied rigorously but it is easier to perform a simulation.

The result, shown in Fig. 8, is surprising. The variance of both α and β increases almost linearly with d . For α the slope of the line is about $\frac{1}{3}$, and the average α^3 has much smaller variance than $\alpha(1)$. For β , however, the slope is about $\frac{4}{5}$. The increase is so strong that averaging over different lags does not improve the accuracy of the estimator.

We have no space to extend this discussion, but for historical reasons we should mention the groundbreaking result of Bienaymé (1874). He considered turning points for white noise as model for series of astronomical measurements (Bienaymé 1875). His 1874 paper was one page without proof. On the last page of his second paper, it is noted that J. Bertrand presented a proof on the next session of the French society for mathematics and astronomy. We slightly modify the result and add the statistics of up-steps.

Bienaymé’s theorem Consider a sequence of T independent and identically distributed random variables. The number V of turning points is asymptotically normal with

$$M(V) = \frac{2}{3}(T - 2) \quad \text{and} \quad \text{Var}(V) = \frac{8}{45}(T - 2) + \frac{1}{30}.$$

The number U of up-steps is asymptotically normal with

$$M(U) = \frac{1}{2}(T - 1) \quad \text{and} \quad \text{Var}(U) = \frac{1}{12}(T - 1) + \frac{1}{6}.$$

Proof In an i.i.d. sequence of length m , all permutations of length m are represented with the same probability. This will be the main argument.

We start with up-steps. Note that $U = \sum_{t=1}^{T-1} U_t$ where $U_t = \mathbf{1}_{\{X_t < X_{t+1}\}}$ denotes the indicator function of up-step t . Since $M(U_t) = p_{12} = \frac{1}{2}$ for each t , we get $M(U) = \frac{1}{2}(T - 1)$. The variance $Var(U) = M(U^2) - M(U)^2$ is expressed as the sum

$$\sum_{t=1}^{T-1} \sum_{s=1}^{T-1} (M(U_s U_t) - \frac{1}{4}) = \sum_{t=1}^{T-1} (M(U_t^2) - \frac{1}{4}) + 2 \sum_{t=1}^{T-2} (M(U_t U_{t+1}) - \frac{1}{4}) .$$

We used that $M(U_s U_t) = \frac{1}{4}$ for $|s - t| > 1$ because of independence. Now $M(U_t^2) = \frac{1}{2}$ and $M(U_t U_{t+1}) = p_{123} = \frac{1}{6}$ implies

$$Var(U) = \frac{T - 1}{4} - 2 \cdot \frac{T - 2}{12} = \frac{T + 1}{12} .$$

For turning points, let $V_t = \mathbf{1}_{\{t \text{ is turning point}\}}$ for $t = 2, \dots, T - 1$. Then $M(V_t) = p_{132} + p_{213} + p_{231} + p_{312} = \frac{4}{6}$ implies that $V = \sum_{t=2}^{T-1} V_t$ fulfils $M(V) = \frac{2}{3}(T - 2)$.

To expand $Var(V) = M(V^2) - M(V)^2$ as a sum, we now have to subtract $\frac{4}{9}$ instead of $\frac{1}{4}$, and we can neglect terms with $|s - t| > 2$ for which V_s and V_t are independent. We obtain $M(V_t V_{t+1}) = 10/24$ by looking at Table 1 and realising that 10 of the 24 permutations have two turning points in the middle. In terms of the covariance we get $Cov(V_t, V_{t+1}) = M(V_t V_{t+1}) - \frac{4}{9} = -1/36$.

It takes a little longer to see that exactly 54 of 120 permutations of length 5 have turning points at their second and second-to-last position. This implies that $Cov(V_t, V_{t+2}) = M(V_t V_{t+2}) - \frac{4}{9} = 1/180$. Then

$$Var(V) = (T - 1)Var(V_t) + 2(T - 2)Cov(V_t, V_{t+1}) + 2(T - 3)Cov(V_t, V_{t+2})$$

gives the result after a brief calculation. Asymptotic normality for $T \rightarrow \infty$ should be rigorously formulated in terms of standardized random variables. In Bienaymé’s time it was taken for granted. Today it follows from central limit theorems for k -dependent or strongly mixing sequences of random variables. □

Note that the variances here are even smaller than for the binomial distribution in Proposition 3, because of negative correlations. This often happens for the variance of pattern frequencies p_π since many patterns, for instance 132, are not compatible with their shifted pattern. Moreover, Bienaymé’s theorem holds for any lag d , in contrast to Proposition 3. But now let us turn to change points.

9 Change points with respect to mean and order structure

Certainly there is no unique way to determine change points in a financial time series. We start by looking for changes in mean, even though from a conceptual viewpoint,

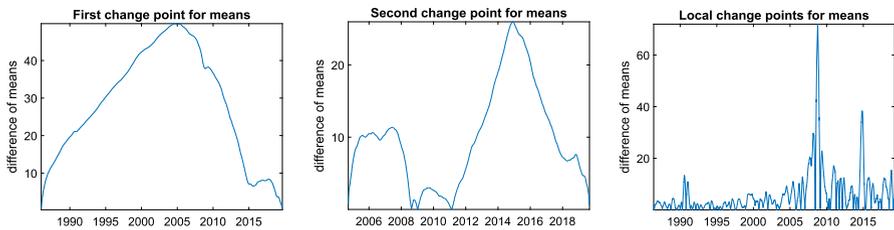


Fig. 9 First and second change point, and local change points in the mean for the oil prices

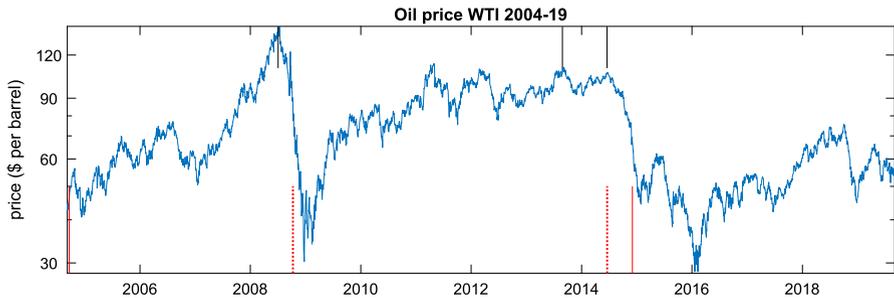


Fig. 10 The oil price series from 2004 to 2019. Calculated change points in mean are indicated at the bottom, local change points by dotted line. Change points in order are indicated at the top

the underlying process is integrated rather than stationary. A classical method takes the time point k where the difference of the means $m_k = \frac{1}{k} \sum_{t=1}^k x_t$ and $\tilde{m}_k = \frac{1}{T-k} \sum_{t=k+1}^T x_t$ assumes a maximum or minimum. Thus we maximize the function

$$f(k) = c_k |m_k - \tilde{m}_k| \quad \text{with } c_k = 2\sqrt{k(T-k)}/T. \quad (10)$$

The factor c_k normalizes the standard deviation of the difference of means, as in a two-sample t -test. (We did not estimate the standard deviation since that would require information on autocorrelation.) We put the constant 2 into c_k so that the factor is 1 for all k in the middle of the interval, and $f(k)$ can be interpreted as a difference.

As seen in Fig. 9, the function f for the WTI series is rather smooth and has a unique maximum point k_1 in the strictly increasing part in the middle of the series in Fig. 5, in late 2004. All x_t with $t < k_1$ are smaller, and almost all x_t with $t > k_1$ are greater than the price at k_1 . The average difference is 50 \$. This is a clear solution to the optimization problem. However, it is not a point where change happens.

When the method is applied to the time period between k_1 and T , we again obtain a clear maximum at the end of 2014. It separates the time of very high prices before 2015 from medium prices we have had since then. Situated on the middle of a rapidly decreasing branch, this is again not a point where change happens.

We can also detect change in the mean locally, by comparing the last 100 points before x_k with the mean of the next 100 points. As shown in Fig. 10, this results in two clear maximum points. One is in the middle of the 2008 period of rapid decrease.

The other one at the beginning of decrease in 2014 seems a real change point. A lot of smaller maxima indicate smaller periods of increase and decrease.

Now we turn to order structure. Let $q^k = (q_1, \dots, q_{24})$ denote the vector of order 4 pattern frequencies for the time interval $[1, k]$, and \tilde{q}^k the respective vector for the time period $[k + 1, T]$. Let $\|v\|$ denote the Euclidean norm of a vector v . Similar as in Sinn et al. (2013), we consider the function

$$g(k) = c_k \|q^k - \tilde{q}^k\| \quad (11)$$

where the constant c_k is the same as in (10). It serves to diminish the bad estimates of frequencies for k near the marginal values 1 and T , and can be justified by the $1/\sqrt{T}$ magnitude of the standard deviation of the pattern frequencies discussed above. Maxima of g are taken as change points for order.

The function $g(k)$ is shown in three versions in the upper panel of Fig. 11. For lag $d = 1$, no clear change points are found. The maximum is near the right margin and turns out to be a random effect too large to be eliminated by c_k . When we take q from a mean of lags, 1 to 3, or 1 to 6, we see clear maxima at the end of 1998, in 2008 and 2013/14. The maximum at the left margin is again a random effect due to high variance of g for small k . On the right, the maximum found for $d = 1$ has disappeared.

When we compare with the time series in Fig. 5, we see that the peaks in July 2008 and July 2014 were included in our ad hoc segmentation. The peak in December 1998 is also a convincing change point. Probably it is a better choice than our first segmentation point in October 2001. And the peak in August 2013 is also reasonable. It could be taken as an alternative to July 2014, as can be seen in Fig. 10. Moreover, the two means of lags give almost identical results, with few days difference in their maximum points. Distances of means over 1 to 6 are on a larger level than those from means over 1 to 3, because of larger variance discussed at Fig. 8.

In the lower panel of Fig. 11, the same results are shown for patterns of length 3. Again, $d = 1$ is not a good option. It was surprising that for means over lags 1 to 3 and 1 to 6 we get the same structure of maxima as for length 4. This indicates that with respect to change points the six patterns of length 3 contain all relevant information.

Can we simplify even further? There are three objectives: the function g should become less erratic, more like the function f for the mean. The criterion should be transparent, interpretable, as it is for the mean. And the estimates of structure before and after should be fairly accurate even for small k or $T - k$. This was not the case for (11) where 1000 values would be minimal for a reasonable estimate of all 24 pattern frequencies. Even with c_k there are marginal effects in Fig. 11.

In the upper panel of Fig. 12. we tested the function $h(k) = c_K (H_k - \tilde{H}_k)$ where H, \tilde{H} denote permutation entropy of order 3 before and after k , respectively. The maximum in July 1998 is a realistic change point, but it is not convincing. In the middle panel we used conditional permutation entropies recommended by Unakafov and Keller (2018). They work well for time series from certain dynamical systems with a Markov structure. We have the same function $h(k)$ with $H = \sum_{i=1}^4 -p^i \log p^i + p \log p + (1 - p) \log 1 - p$ where $p = p_{12}, p^1 = p_{123}, p^2 = p_{132} + p_{231}, p^3 = p_{213} + p_{312}$ and $p^4 = p_{321}$. The function is rather smooth but there is no clear maximum. Our data do not include a Markov structure. In the lower panel we used

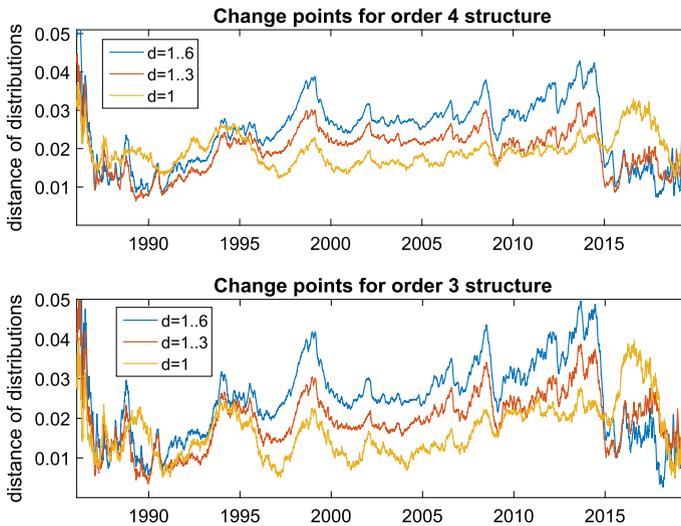


Fig. 11 Distance of pattern distributions before and after time k . Means of lags give better results, while patterns of length 3 or 4 lead to the same change points

turning rates, $h(k) = c_k(\alpha_k - \tilde{\alpha}_k)$. As noted in Sect. 2 they worked so well for classifying sleep brain data. Here they give almost the same result as permutation entropy. All three functions seem not appropriate for segmenting our financial data.

10 Change points with respect to up-down balance

Our last trial is the function β . It will give the best results. We look for maxima and minima of $h(k) = c_k(\beta_k - \tilde{\beta}_k)$ where β_k and $\tilde{\beta}_k$ are the values of β on $[1, k]$ and $[k + 1, T]$, respectively. Note that by Proposition 3, c_k is exactly the normalizing constant for β and α , at least for lag 1. Figure 13 shows that lag 1 alone is sufficient only from 1992 to 2002. The other two estimators of β give very similar functions. Their main maximum and minimum points are very near to the maxima for the order structure in Fig. 11. The first change point k_1 has to be taken in August 2013. Our ad hoc point in July 2014 would be a possible alternative choice.

As in the approach with means, we can now look for a second change point k_2 in the longer subinterval, in this case $[1, k_1]$. Figure 14 shows that the second change point is unique when we exclude margins. It is a minimum, which means that prices start to rise there, and can already be seen in Fig. 13. Here it comes one month later than in Fig. 11, in early February 1999. When we look for a third change point in the interval $[k_1, k_2]$ we again get a unique solution which is already seen as maximum in Fig. 14. This point in July 2008 differs only by one day from our ad hoc point with the highest oil price. Thus β produces a meaningful segmentation which is similar to, and better than, our ad hoc segmentation.

One can also search for local change points, comparing β on the intervals $[k - m + 1, k]$ and $[k + 1, k + m]$. The situation is similar as for means: the local search

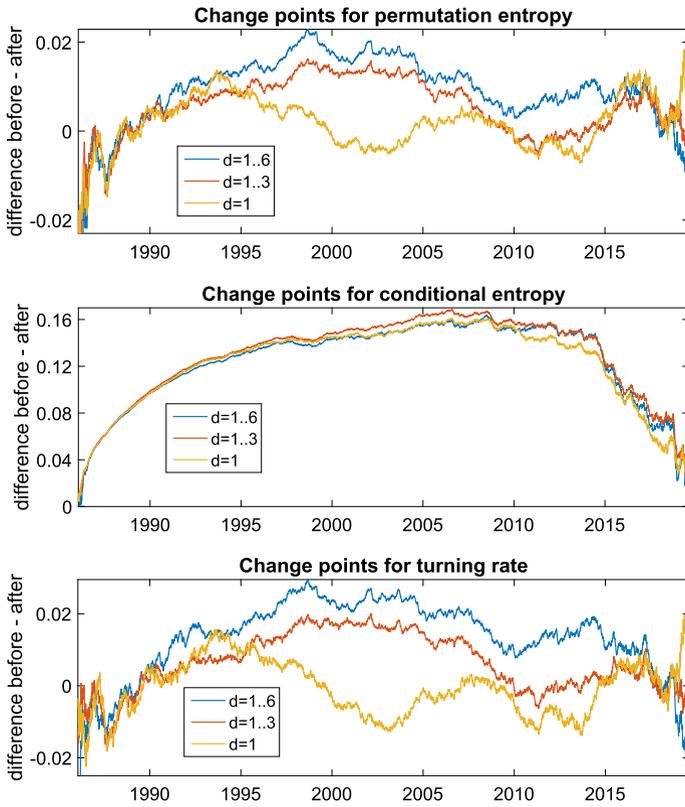


Fig. 12 Difference before and after k of order 3 permutation entropies, conditional entropies recommended by Unakafov and Keller (2018), and turning rates. These methods seem not appropriate for segmentation of financial data

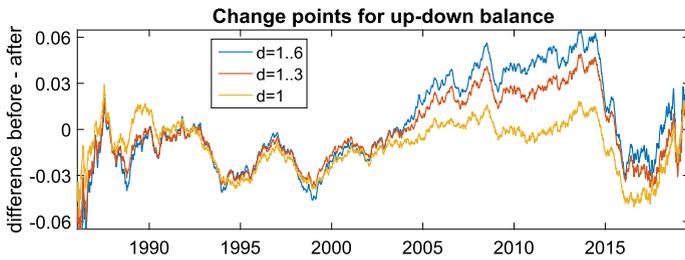


Fig. 13 Difference of β -values before and after k . Up-down balance seems the most simple and most relevant parameter for change points in financial time series

gives a lot of maxima and minima, which depend very much on the scale parameter m . Figure 14 shows the result for $m = 150$. Two large maxima in 2008 and 2014 are similar to those in Fig. 9. Nevertheless, local search offers too many choices.

For financial data, global search with β seems the best order method for finding change points. It provides more meaningful results than the approach with means.

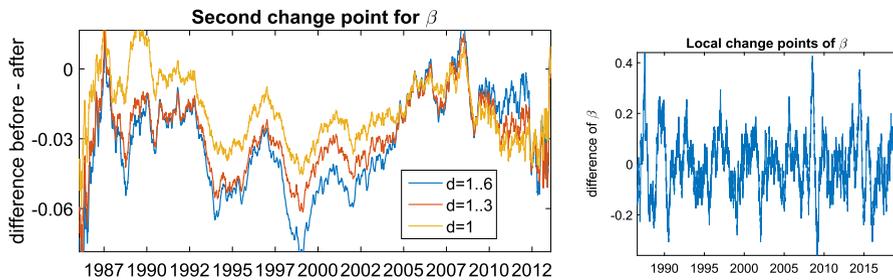


Fig. 14 Second change point for β , and local change points for β the oil prices

What about significance of $h(k_1)$? In view of Proposition 3, we can think about random walk statistics to provide p -values. Calculation of β on intervals $[1, k]$ can be seen as a simple symmetric random walk b_1, \dots, b_T with $T - 1$ time steps. The initial value is $b_1 = 0$, and the terminal value b_T is the observed total number of up-steps. The probability p_c that the walk will hit the lines $y = \pm c$ can be expressed by binomial coefficients.

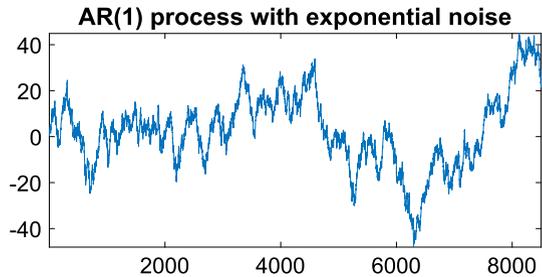
This exciting idea has several shortcomings. First, we need the difference between relative frequencies, normalized by c_k , while random walk concerns absolute values. Next, it would work only for lag 1, which from a practical viewpoint is not the best option for change point detection in financial time series. The most serious problem, however, is that the question is wrong. It is easy to define changes in the mean or dynamics of model series. But do we really expect similar dynamical changes in a given real-world time series?

For β as mean over lags 1 to 3, we have the maximum value $h(k_1) = 0.548$ in Fig. 13. We checked the significance of this change point by simulation, with Brownian motion as null model. The maximum value of $|h(k)| = c_k \cdot |\beta_k - \tilde{\beta}_k|$ was determined for 1000 trajectories of BM. It turned out that for 422 simulations, the maximum was larger than 0.548. Almost every second realization of BM had a change point with larger value $|h(k_1)|$ than our data.

This is alright! BM is not stationary. A time series taken from BM needs segmentation, like the data. Any reasonable method should detect change points in BM. As a null model, we better take a stationary series which by definition has no change points. We first took an AR(1) model $x_t = 0.99x_{t-1} + z_t$ with i.i.d. Gaussian random numbers z_t . We found that only 2 from 1000 realizations have larger $|h(k_1)|$ than our data. So for stationary series, we get significantly smaller values in the search for change points.

We checked this with the same AR(1) process with exponential noise, $z_t = 1 - e_t$, where e_t has an exponential distribution with mean 1. These random numbers have their heavy tail on the left, so the process will make downwards excursions and only slowly climb up. As a result, $\beta(1) \approx 0.25$, and all $\beta(d)$ with $d \leq 6$ are still greater 0.1. Nevertheless, only two from 1000 simulations gave a maximum larger 0.548. Then we realized that the excursions are still too short, and increased the AR(1) factor to 0.998. A typical realization with maximum h -value 0.32 is shown in Fig. 15. For this process, 92 from 1000 simulations had maxima greater 0.548.

Fig. 15 Sample from an AR(1) with exponential noise, with positive $\beta(d)$ for small d . Although the process is stationary by definition, this cannot be seen in our figure, and segmentation may be necessary



These experiments gave the impression that β is a useful parameter for change point detection. It will provide large values of h when we expect them. But this is exploratory research, there is still much work to do. We hope that readers feel encouraged to perform their own studies with order patterns.

Acknowledgements Open Access funding provided by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Amigo JM (2010) Permutation complexity in dynamical systems. Springer series in synergetics. Springer, Berlin
- Amigo J, Keller K, Kurths J (2013) Recent progress in symbolic dynamics and permutation complexity. ten years of permutation entropy. *Eur Phys J Spec Top* 222:247–257
- Amigo JM, Keller K, Unakafova VA (2015) Ordinal symbolic analysis and its application to biomedical recordings. *Philos Trans R Soc A* 373:20140091
- Bandt C (2015) Permutation entropy and order patterns in long time series. In: Rojas I, Pomares H (eds) *Time series analysis and forecasting, contributions to statistics*. Springer, New York
- Bandt C (2017) A new kind of permutation entropy used to classify sleep stages from invisible EEG microstructure. *Entropy* 19(5):197
- Bandt C (2018) Crude eeg parameter provides sleep medicine with well-defined continuous hypnograms [ArXiv:1710.00559](https://arxiv.org/abs/1710.00559)
- Bandt C (2019) Small order patterns in big time series: a practical guide. *Entropy* 21(6):613. <https://doi.org/10.3390/e21060001>
- Bandt C, Pompe B (2001) Permutation entropy: a natural complexity measure for time series. *Phys Rev Lett* 88:174102
- Bandt C, Shiha F (2007) Order patterns in time series. *J Time Series Anal* 28(5):646–665
- Betken A, Buchsteiner J, Dehling H, Munker I, Schnurr A, Woerner J (2019) Ordinal patterns in long-range dependent time series. [ArXiv:1905.11033](https://arxiv.org/abs/1905.11033)
- Bienaymé I (1874) Sur une question de probabilités. *Bull Soc Math France* 2:153–154
- Bienaymé I (1875) Application d'un théorème nouveau au Calcul des probabilités. *Bulletin des sciences mathématiques et astronomiques* 9:219–225

- Bruzzo A, Gesierich B, Santi M, Tassinari C, Birbaumer N, Rubboli G (2008) Permutation entropy to detect vigilance changes and preictal states from scalp EEG in epileptic patients. a preliminary study. *Neurol Sci* 29:3–9
- Chicote B, Irusta U, Alcaraz R, Rieta JJ, Aramendi E, Isasi I, Alonso d, Iburguren K (2016) Application of entropy-based features to predict defibrillation outcome in cardiac arrest. *Entropy* 18:313
- Embrechts P, Maejima M (2002) Selfsimilar processes. Princeton University Press, Princeton, NJ
- Embrechts P, Klüppelberg C, Mikosch T (1997) Modelling extremal events for insurance and finance. Springer, New York
- Ferlazzo E et al (2014) Permutation entropy of scalp EEG: a tool to investigate epilepsies. *Clin Neurophysiol* 125:13–20
- Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov P, Mark R, Mietus J, Moody G, Peng CK, Stanley H (2000) Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* 101(63):e215–e220
- Kuo CE, Liang SF (2011) Automatic stage scoring of single-channel sleep EEG based on multiscale permutation entropy. In: Biomedical circuits and systems conference (BioCAS), IEEE, pp 448–451
- Mandelbrot BB (1997) Fractals and scaling in finance. Springer, New York
- McCullough M, Small M, Iu H, Stemler T (2017) Multiscale ordinal network analysis of human cardiac dynamics. *Philos Trans R Soc A* 375(2096):20160292
- Morabito FC, Labate D, Foresta FL, Bramanti A, Morabito G, Palamara I (2012) Multivariate multi-scale permutation entropy for complexity analysis of Alzheimer's disease EEG. *Entropy* 14:1188–1202
- Nicolaou N, Georgiou J (2012) The use of permutation entropy to characterize sleep encephalograms. *Clin EEG Neurosci* 39:202–209
- Olofsen E, Sleight J, Dahan A (2008) Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect. *Br J Anaesth* 101:810–821
- Ouyang G, Dang C, Richards D, Li X (2010) Ordinal pattern based similarity analysis for eeg recordings. *Clin Neurophysiol* 121:694–703
- Parlitz U, Berg S, Luther S, Schirdewan A, Kurths J, Wessel N (2012) Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics. *Comput Biol Med* 42(3):319–327
- Schnurr A (2014) An ordinal pattern approach to detect and to model leverage effects and dependence structures between financial time series. *Stat Pap* 55(4):919–931
- Schnurr A, Dehling H (2017) Testing for structural breaks via ordinal pattern dependence. *J Am Stat Assoc* 112:706–720
- Sinn M, Keller K (2011) Estimation of ordinal pattern probabilities in Gaussian processes with stationary increments. *Comput Stat Data Anal* 55:1781–1790
- Sinn M, Keller K, Chen B (2013) Segmentation and classification of time series using ordinal pattern distributions. *Eur Phys J Spec Top* 222:587–598
- Terzano M, Parrino L, Sherieri A, Chervin R, Chokroverty S, Guilleminault C, Hirshkowitz M, Mahowald M, Moldofsky H, Rosa A, Thomas R, Walters A (2001) Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (cap) in human sleep. *Sleep Med* 2(6):537–553
- Unakafov AM, Keller K (2018) Change-point detection using the conditional entropy of ordinal patterns. *Entropy* 20:709
- Zanin M, Zunino L, Rosso O, Papo D (2012) Permutation entropy and its main biomedical and econophysics applications: a review. *Entropy* 14:1553–1577
- Zunino L, Zanin M, Tabak B, Pérez D, Rosso O (2009) Forbidden patterns, permutation entropy and stock market inefficiency. *Physica A* 388:28543–2864
- Zunino L, Bariviera A, Guercio M, Martinez L, Rosso O (2012) On the efficiency of sovereign bond markets. *Physica A* 391:4342–4349
- Zunino L, Bariviera A, Guercio M, Martinez L, Rosso O (2016) Monitoring the informational efficiency of European corporate bond markets with dynamical permutation min-entropy. *Physica A* 456:1–9