A multi-scale symmetry analysis of daily financial time series using uninterrupted trends returns

Carlos Manuel Rodríguez Martínez
David Hernandez Enríquez
Advisor: Dr. Alejandro Raúl Hernández Montoya

Universidad Veracruzana

10 de julio de 2019

Acknowledgments

- Dr. Alejandro Raúl Hernandez Montoya. Universidad Veracruzana.
- Dr. Hector Conoronel Brizio. Universidad Veracruzana.

Introduction

Publicly available information

Prices, volume, dividends, beta, etc.

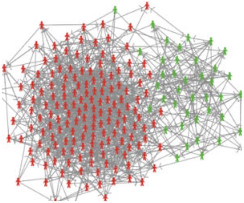
Intel Corporation (INTC)

Best Bid/Ask	$ 29.96 / $ 29.97
1 Year Target	38
Today's High /Low	$ 30.50 / $ 29.45
Share Volume	52,122,067
50 Day Avg. Daily Volume	22,872,515
Previous Close	$ 32.74
52 Week High/Low	$ 37.03 / $ 24.87
Market cap	$ 141,428,430,000
P/E Ratio	12.81
Forward P/E(1y)	13.76
Earnings Per Share (EPS)	$ 2.34
Annualized dividend	$ 0.96
Ex Dividend Date	Nov. 4, 2015
Dividend Payment Date	Dec. 1, 2015
Current Yield	2.93 %
Beta	1.11
NASDAQ Official Open Price	$ 29.63
Date of Open Price	Jan. 15, 2016
NASDAQ Official Close Price	$ 32.74
Date of Close Price	Jan. 14, 2016
Community Sentiment	Bullish

What is relevant to price

- Interaction of many agents that perform buy/sell operations.
- Every agent has access to the same public information to make an informed decision on every operation. They are rational.
- The direct consequence is that current price already contains all the knowledge of future forecasts.

$$E\left[P_{t+1} | P_0,\ P_1, \ldots, P_t\right] = P_t.$$



Efficient market hypothesis (EMH)

- **Bachelier (1900)** proposes that since in a voluntary stock exchange between two agents, each of them believes them to be the winning party. If the game is fair (Martingale) it must behave like a random walk. His work was not well received.
- **Samuelson (1965)** proposes the efficient market hypothesis. If investors were able to incorporate future events in the current price, they would have done so, incorporating their knowledge in the current price.
- **Fama (1995)** finds empirical evidence that markets behave like a random walk.
- The EMH requires agents to be rational.

An historical example

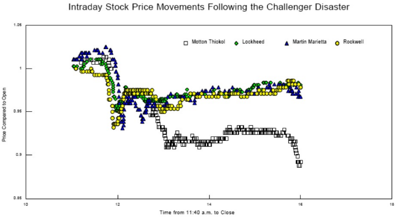
- Challenger space shuttle. Four contractors involved: Lockheed, Martin Marietta, Morton Thiokol y Rockwell.
- After 5 months of investigation it was found that the misbehaving component came from Morton Thiokol.
- Markets declared their verdict right away.

Efficient Markets - 5

Efficient Markets – Challenger Example

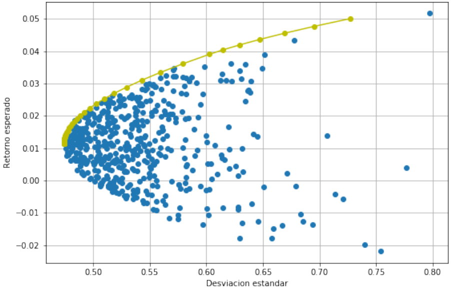
Source: Michael T. Maloneya and J. Harold Mulheri, The complexity of price discovery in an efficient market: the stock market reaction to the Challenger crash, Journal of Corporate Finance. Volume 9, Issue 4 , September 2003, Pages 453-479

- Here are the stock prices for several major suppliers to NASA (for the day)



If I can't win there is no point

- **Markowitz (1952)** optimal portfolio theory tell us that we can construct a portfolio such that the relation between expected return and risk is optimal.



- Useful for long term investment.
- EMH implies the reliability of predictive markets.

There are winners however...

- The existence of agents that can reliably generate profits like Warren Buffet, Peter Lynch and George Soros, implies the possibility that some agents can sometimes predict better than others future outcomes or even affect the market themselves.
- Humans have cognitive biases that systematically makes us averse to lose and have high hopes of winning. We are not always rational.

Motivation

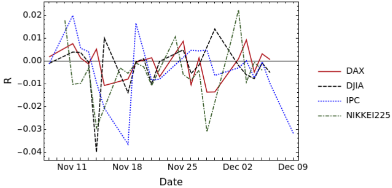
How can we measure if the market really behaves like a random walk?

Background

Definitions

- We define the logarithmic returns as

$$R(t, \Delta t) = \log P_{t+\Delta t} - \log P_t.$$



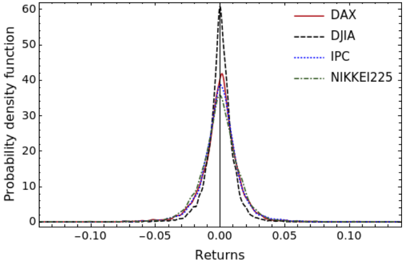
- We can express the efficient market hypothesis in terms of the returns

$$E\left[R_{t+1} | R_0,\ R_1, \ldots, R_t\right] = 0.$$

- From this definition of the returns emerge the *stylized facts*, that are statistical properties common to the returns of most markets.

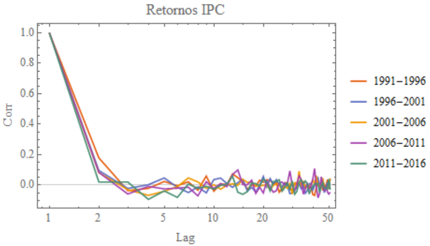
Stylized facts I

- Win / Loss assymetry. Symmetry is strongly related to the EMH.

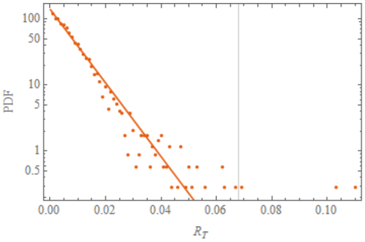


- No auto-correlation on the returns

Stylized facts II

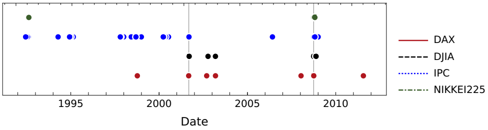


- Returns decay as a power law.



Stylized facts III

- Some events extremely rare (according to the power law) still happen sometimes. We identify these extreme events as the Dragon Kings described by Sornette. Usually these extreme values signal a big transition in the market like an economic crisis.

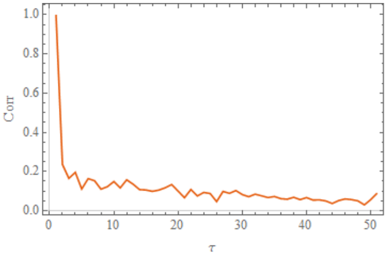


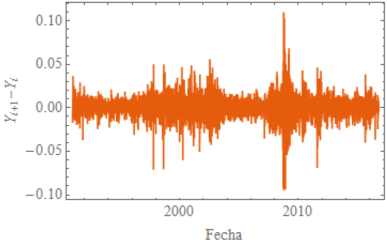
- Volatility clustering

$$C_2(\tau) = \mathrm{Corr}(R(t+\tau,\Delta t)^2,\ R(t,\Delta t)^2).$$

That is, big changes tend to be followed by big changes of any sign, and small changes tend to be followed also by small changes of any sign.

Stylized facts IV





H.F. Coronel-Brizio, et al. (2007) Assessing symmetry of financial returns series. Physica A 383.

A test for symmetry

- Based on the T_n statistic developed by Einmahl and McKeague.
- The null hypothesis is that there is a symmetry point around c, that is $H_0 : F(c - x) = 1 - F(x - c)$, for every $x > 0$ in the sample $X_1, \ldots, X_n$. F is a cumulative distribution function.

Einmahl, McKeague (2003) Empirical likelihood based hypothesis testing I

Many of the papers cited above consider the case of a known point of symmetry and use a Cramér–von Mises type test statistic. We also assume that the point of symmetry is known, so without loss of generality it is assumed to be zero. Let $X_1, \ldots, X_n$ be i.i.d. with continuous distribution function F. The null hypothesis of symmetry about zero is

$$H_0 : F(-x) = 1 - F(x-), \qquad \text{for all } x > 0.$$

The local likelihood ratio statistic is defined by

$$R(x) = \frac{\sup\{L(\tilde{F}) : \tilde{F}(-x) = 1 - \tilde{F}(x-)\}}{\sup\{L(\tilde{F})\}}, \qquad x > 0.$$

As in the Introduction, the unrestricted likelihood in the denominator is maximized by setting $\tilde{F} = F_n$, the empirical distribution function. The supremum in the numerator can be found by treating $\tilde{F}$ as a function of $0 \leqslant p \leqslant 1$, where $\tilde{F}$ puts mass $p/2$ on the interval $(-\infty, -x]$, mass $p/2$ on $[x, \infty)$ and mass $1 - p$ on $(-x, x)$, with those masses divided equally among

Einmahl, McKeague (2003) Empirical likelihood based hypothesis testing II

the observations in the respective intervals. That is, the masses on the individual observations in the respective intervals are given by

$$\frac{p/2}{n\hat{p}_1}, \frac{p/2}{n\hat{p}_2}, \frac{1-p}{n(1-\hat{p})},$$

where $\hat{p} = \hat{p}_1 + \hat{p}_2$, $\hat{p}_1 = F_n(-x)$ and $\hat{p}_2 = 1 - F_n(x-)$. The numerator of $R(x)$ is therefore the maximal value of

$$\left(\frac{p/2}{n\hat{p}_1}\right)^{n\hat{p}_1} \left(\frac{p/2}{n\hat{p}_2}\right)^{n\hat{p}_2} \left(\frac{1-p}{n(1-\hat{p})}\right)^{n(1-\hat{p})},$$

which is easily seen to be attained at $p = \hat{p}$. We thus obtain

$$\begin{aligned} \log R(x) &= n\hat{p}_1 \log \frac{\hat{p}}{2\hat{p}_1} + n\hat{p}_2 \log \frac{\hat{p}}{2\hat{p}_2} \\ &= nF_n(-x)\log \frac{F_n(-x) + 1 - F_n(x-)}{2F_n(-x)} \\ &\quad + n(1 - F_n(x-))\log \frac{F_n(-x) + 1 - F_n(x-)}{2(1 - F_n(x-))}, \end{aligned} \tag{2.1}$$

H.F. Coronel-Brizio, et al. (2007) Assessing symmetry of financial returns series. Physica A 383.

- The statistic T_n around a test point c is

$$T_n = -\frac{2}{n}\sum_{i=1}^{n}\log R(|X_i|).$$

where

$$\log R(x) = nF_n(c-x)\log\frac{F_n(c-x)+1-F_n(x-c)}{2F_n(c-x)} + n(1-F_n(x-c))\log\frac{F_n(c-x)+1-F_n(x-c)}{2(1-F_n(x-c))}.$$

and F_n is the cumulative empirical distribution function.

H.F. Coronel-Brizio, et al. (2007) Assessing symmetry of financial returns series. Physica A 383.

- We can test for c in an interval $[C_{min}, C_{max}]$, for example

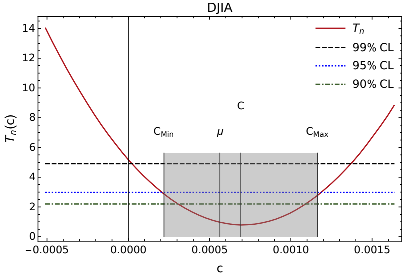


Figura: T_n symmetry test evaluated on the returns of the DJIA index.

- We can affirm with 95 % certainty that the distribution is symmetrical in a interval between 0.0002 and 0.001.

The problem of scale

- The time lag in which the returns are defined are particularly relevant when measuring symmetry.

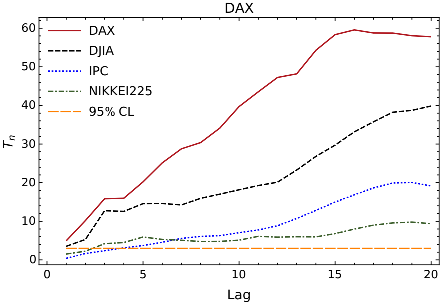


Figura: $T_n(0)$ with returns of different lags.

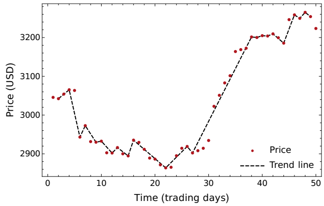
- Artifacts are introduced with this methodology.

Definition of trend returns

- Multiscale return definition. Given a financial time series $P_1, \ldots, P_n$, we define an "uninterrupted trend" of duration k, a succession of $k+1$ consecutive values of the given time series where each value is greater than the preceding one.
- We define their corresponding trend return as:

$$TRet_m^k := \log(P_{m+k}) - \log(P_m) \tag{1}$$

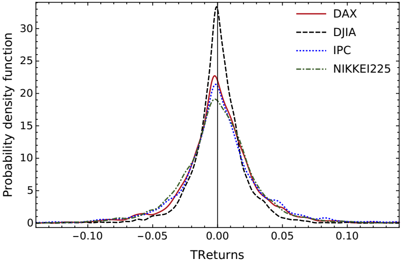
Where m indexes the different trends and k indicates the duration in days of m-Th trend.

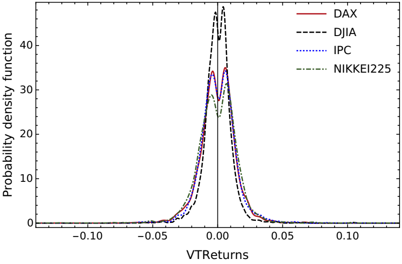


Trend returns distribution

Analogously we can define the *trend velocity returns* as the rate in which the trend returns change.

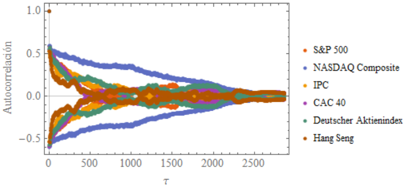
$$TVRet_m^k := \frac{\log(P_{m+k}) - \log(P_m)}{k} \quad (2)$$



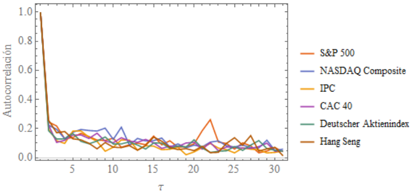


This definition also comes with its stylized facts I

- Volatility clustering is present in trend returns R_m^k



(a) Trend returns autocorrelation. Not very informative.



(b) Quadratic trend returns autocorrelation.

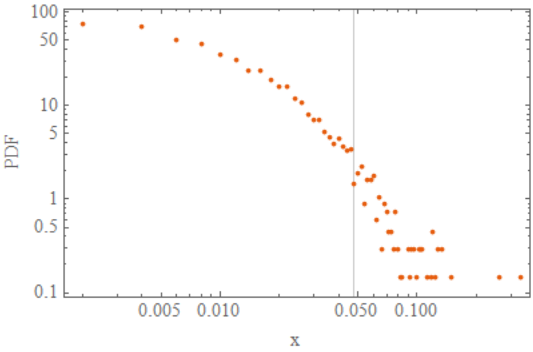
- So trend returns still can be decomposed as

$$R_T(t) = \sigma_R(t)\epsilon(t),$$

where $\epsilon(t)$ is gaussian noise and $\sigma(t)$ the volatility.

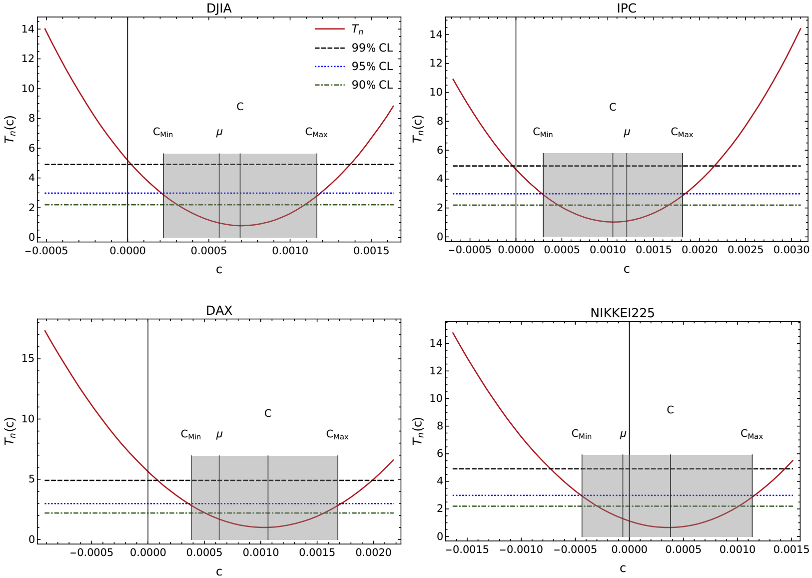
This definition also comes with its stylized facts II

H.F. Coronel-Brizio, A.R. Hernandez-Montoya (2005). On fitting the Pareto-Levy distribution to stock market index data: selecting a suitable cut off value. Physica A 354 437-449. Describes a procedure to select the cutoff point where returns start behaving as a power law.

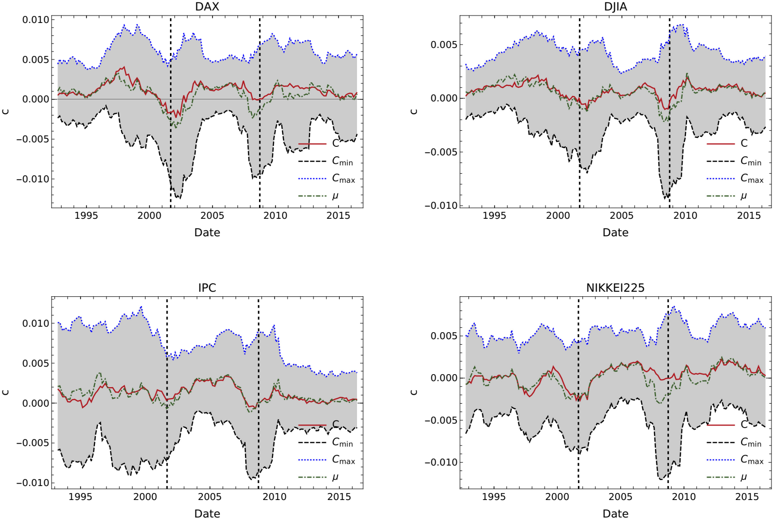


Symmetry: Results

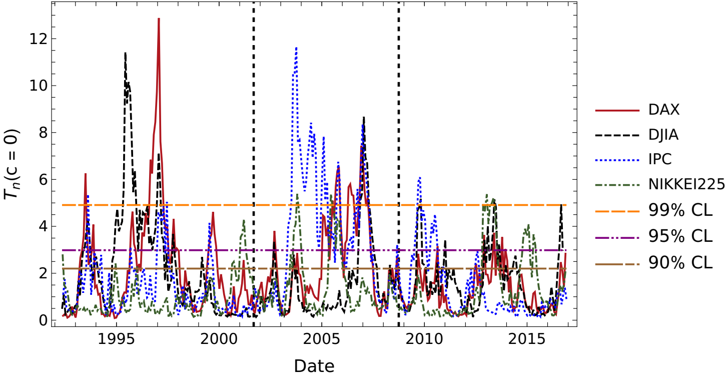
Symmetry tests: Trend returns



Time evolution of the most plausible point of symmetry: Trend returns

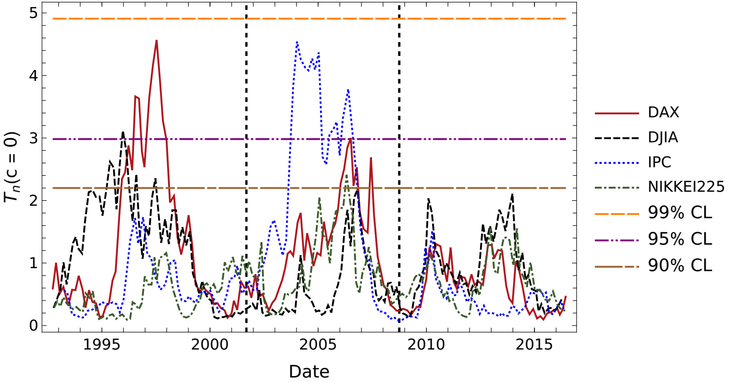


Symmetry around zero: Simple returns



Symmetry around $c = 0$ for simple returns

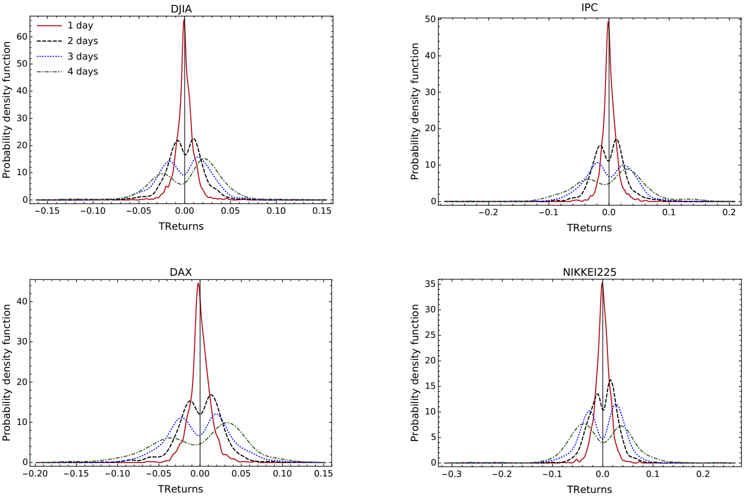
Symmetry around zero: Trend returns



Symmetry around $c = 0$ for trend returns

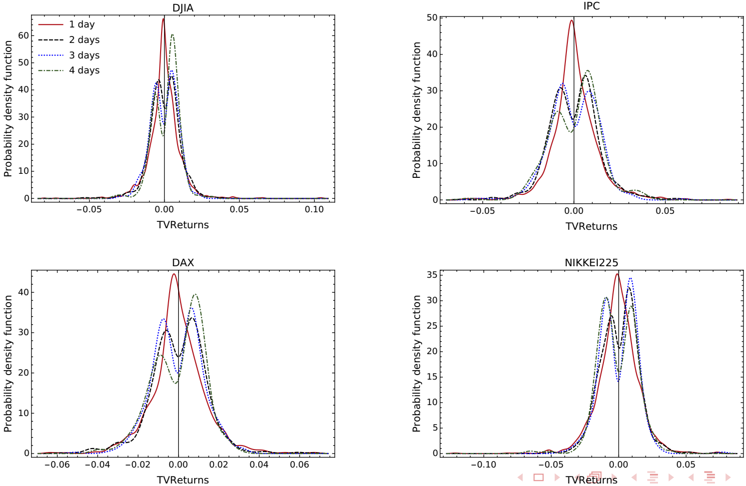
TRet and TVRet signals

The distributions of TRets and TVRets can be separated by run length by days.



TRet and TVRet signals

The shape of distributions of $k > 1$ are more visible in TVRets because of the scaling done when dividing the TRets by run duration.



Distribution of trends

Model

- Among all the possible martingale or sub-martingale models that can describe price fluctuations, the geometric random walk is the simplest one.
- At each step there are two possible outcomes: the index either increases or does not increase. In an efficient market, the expected future price depends only on information about the current price, not on its previous history.
- It should be impossible to predict the expected direction of a future price change given the history of the price process. We have

$$\mathbb{E}(S(t+\Delta t) - S(t)|\mathcal{F}_t) = 0; \tag{3}$$

if we consider the sign of the price change $Y(t, \Delta t) = \mathrm{sign}(S(t+\Delta t) - S(t))$, which coincides with the sign of returns, we accordingly have

$$\mathbb{E}(Y(t, \Delta t)) = 0. \tag{4}$$

Model II

- If the price follows a geometric random walk, then the series of price-change signs can be modeled as a Bernoulli process.+ Let S_0 be the initial price. The price at time t will be given by

$$S(t) = S_0 \prod_{i=1}^{t} Q_i \tag{5}$$

where Q_i are independent and identically distributed random variables following a log-normal distribution with parameters μ and σ.

- These two parameters come from the corresponding normal distribution for log-returns. As a direct consequence of the EMH we have

$$\mathbb{E}(Q) = 1 + r_F, \tag{6}$$

and for a log-normal distributed random variable, we have also

$$\mathbb{E}(Q) = \mathrm{e}^{\mu}\mathrm{e}^{\sigma^2/2}. \tag{7}$$

Model III

- This leads to a dependence between the two parameters

$$\mu = \log(1 + r_F) - \frac{\sigma^2}{2}. \tag{8}$$

- Starting from the cumulative distribution function for a log-normal random variable

$$F_Q(u) = \mathbb{P}(Q \leq u) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{\log(u) - \mu}{\sqrt{2\sigma^2}}\right), \tag{9}$$

the probability of a negative sign would be given by

$$q = F_Q(1) = \mathbb{P}(Q \leq 1) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{\sigma}{2\sqrt{2}} - \frac{\log(1 + r_F)}{\sigma\sqrt{2}}\right), \tag{10}$$

which yields $q = 1/2$ for $r_F = \mathrm{e}^{\sigma^2/2} - 1$.

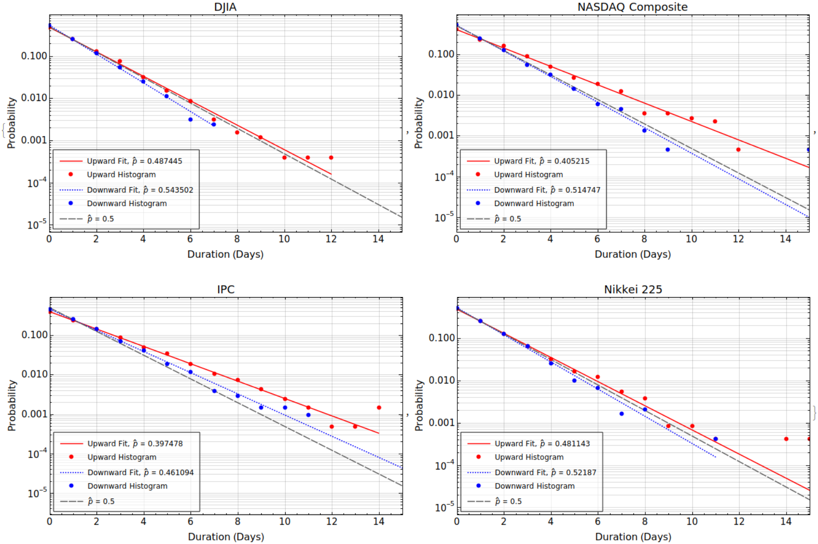
Model IV

- It becomes natural to use the biased Bernoulli process as the null hypothesis for the time series of signs. It is well known that the distribution of the number k of failures needed to get one success for a Bernoulli process with success probability $p = 1 - q$ is the geometric distribution $\mathcal{G}(p)$; the number of failures N is given by

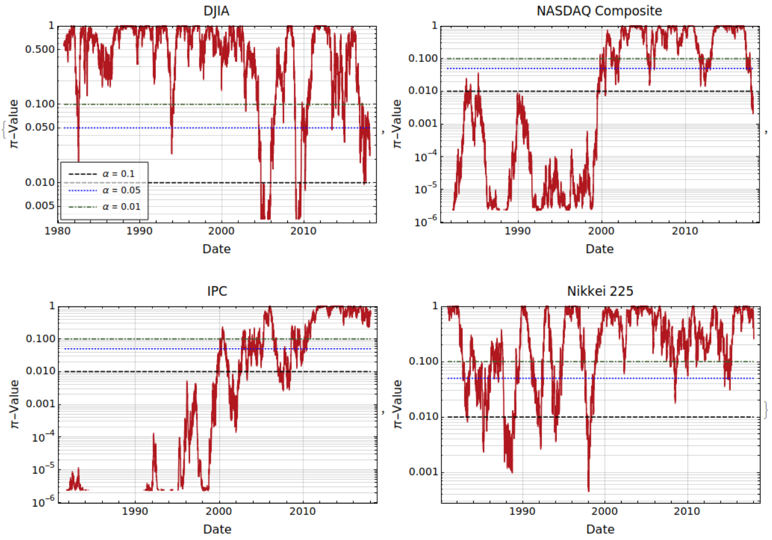
$$P(k) = \mathbb{P}(N = k) = p(1 - p)^k = pq^k. \tag{11}$$

- In order to compare the observed and expected distributions of trend durations, the Anderson-Darling test was used. The Anderson-Darling test was found to be the most suitable for this purpose because it places more weight on the tails of a distribution than other goodness of fit tests.

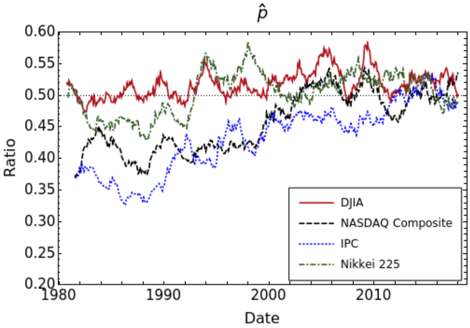
Distribution of runs



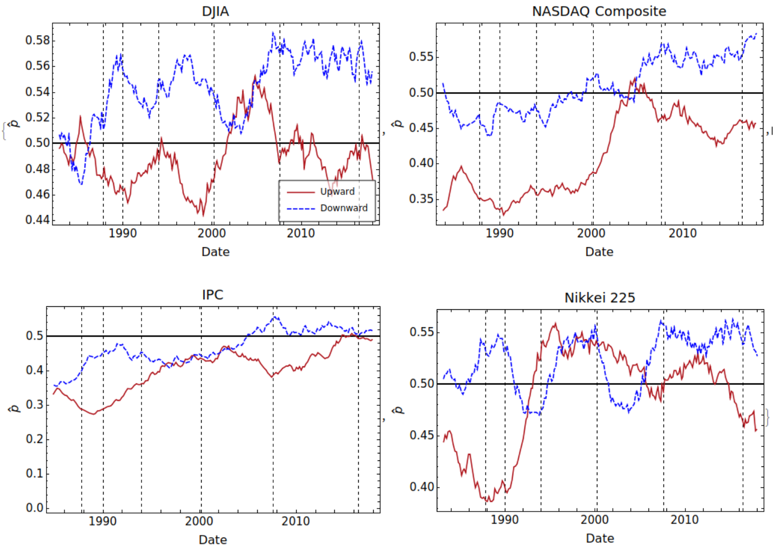
Test the data with a Geometric Distribution $p = 0.5$



Estimation of $\hat{p}$ for the runs with a geometric distribution

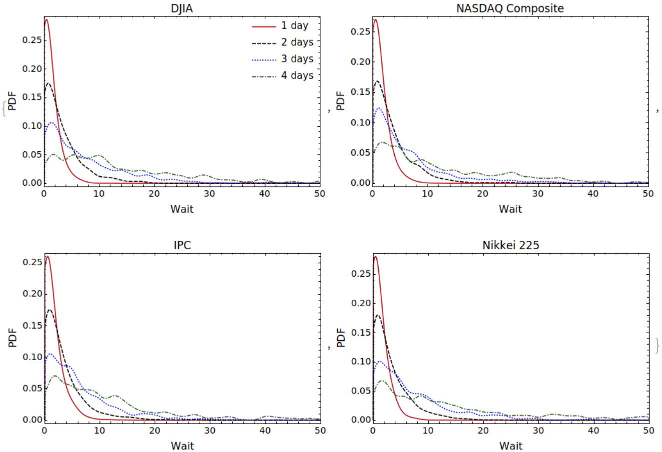


Estimation of $\hat{p}$ for positive/negative runs with a geometric distribution

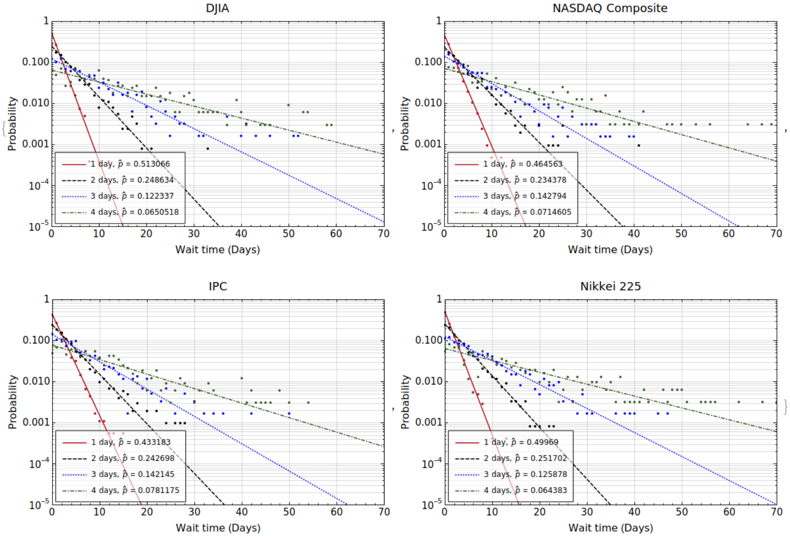


Waiting time distribution

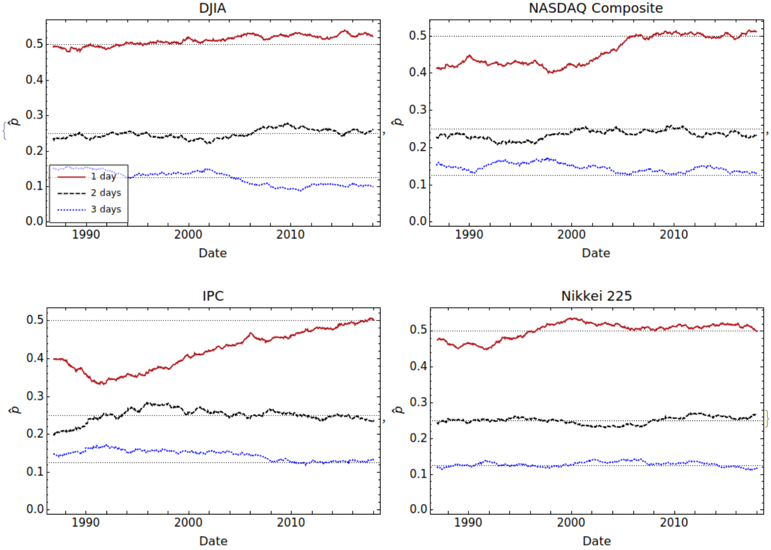
- In order to investigate the cause of the deviation from the random walk we investigated the distribution of waiting times, which are also expected to distribute as a GeometricDistribution.



Total waiting time parameter estimation



Waiting time parameter estimation in time



Conclusion

- We can measure symmetry of the returns using the T_n statistic. We also defined the trend returns in order to select a scale independent observable.
- The most plausible symmetry point is not always correlated with the average.
- Symmetry is not constant. Particularly symmetry around $c = 0$ tends to be plausible in dates closer to extreme events.
- Might be interesting to investigate the dynamics of the run signals.

Thank you all for your attention.

Bibliography I