Accepted Manuscript

A top–bottom price approach to understanding financial fluctuations

Miguel A. Rivera-Castro, José G.V. Miranda, Ernesto P. Borges,
Daniel O. Cajueiro, Roberto F.S. Andrade

PII: S0378-4371(11)00857-0
DOI: 10.1016/j.physa.2011.11.022
Reference: PHYSA 13528

To appear in: Physica A

Received date: 13 July 2011
Revised date: 12 October 2011


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
A top–bottom price approach to understanding financial fluctuations

Miguel A. Rivera-Castro, José G. V. Miranda, Ernesto P. Borges, Daniel O. Cajueiro, Roberto F. S. Andrade

*Instituto de Física, Universidade Federal da Bahia, BA 40210-340, Brazil.
Department of Economics – University of Brasilia, DF, Brazil 70910-900
National Institute of Science and Technology for Complex Systems, Brazil

Abstract

The presence of sequences of top and bottom (TB) events in financial series is investigated for the purpose of characterizing such switching points. They clearly mark a change in the trend of rising or falling prices of assets to the opposite tendency, are of crucial importance for the players’ decision and also for the market stability. Previous attempts to characterize switching points have been based on the behavior of the volatility and on the definition of microtrends. The approach used herein is based on the smoothing of the original data with a Gaussian kernel. The events are identified by the magnitude of the difference of the extreme prices, by the time lag between the corresponding events (waiting time), and by the time interval between events with a minimal magnitude (return time). Results from the analysis of the inter day Dow Jones Industrial Average index (DJIA) from 1928 through 2011 are discussed. $q$-Gaussian functions with power law tails are found to

*Corresponding author

Email addresses: marc@ufba.br (Miguel A. Rivera-Castro), vivas@ufba.br (José G. V. Miranda), ernesto@ufba.br (Ernesto P. Borges), danielcajueiro@unb.br (Daniel O. Cajueiro), randrade@ufba.br (Roberto F. S. Andrade)

Preprint submitted to Physica A October 12, 2011
provide a very accurate description of a class of measures obtained from the series statistics.

1. Introduction

It is well known that the understanding of financial fluctuations is fundamental for appropriate investment management. While financial fluctuations may represent a source for great gains, they may also be very harmful for investors, if their financial portfolios are not suitably protected. In this context, several works have tried to understand the dynamics of financial and economic fluctuations [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. A subject of particular interest for this work has been presented recently [8], where it was proposed that financial fluctuations may be studied through the so-called return intervals approach, where the return interval is defined by time interval between two consecutive volatilities above a given threshold.

In this work we call the attention to the fact that another interesting way to study financial fluctuations is based on the top-bottom (TB) price approach. Top and bottom prices represent switching points in financial series that signalize changes of expectations. While a bottom price means that the asset is being sold by a too low value, a top price means that the asset is being sold by a too high value. In this context, two entities are significant, namely the so-called the TB-return and the TB-interval. We define here the TB-return as the absolute value of the difference between consecutive top and bottom prices or bottom and top prices of price time series. We also define the TB-interval as the time interval between the events of consecutive top and bottom or bottom and top.
Based on these measures, we proceeded with a statistical characterization of the events as function of their magnitudes. Our results show that the occurrence of TB-returns and TB-intervals follow quite distinct patterns. Furthermore, in the search for a more comprehensive analysis, we looked for the existence of memory effects in the series. Long-range analysis in the time series is presented based on results for TB return-intervals (along some already quoted ideas [8]), as well the behavior of the correlation function between TB-returns and TB-intervals. Both analyses of memory effects identify different behaviors for the two measures, which can be correlated with the probability distribution patterns.

The advantage of this methodology is that tops and bottoms of financial time series are important pieces of information for several investors. In particular, several patterns of technical analysis are based on information that is extracted from the relative positions of tops and bottoms in price time series [11]. Furthermore, one should note that TB returns and TB intervals are strongly related and provide important information. The TB return measures the maximal amount of money that an investor can make/lose in a given TB interval when the price of the asset rises/falls.

While it is very easy to determine tops (maximums) and bottoms (minimums) in smooth continuous functions, a procedure is necessary to do that in financial time series. Therefore, in order to determine tops and bottoms in financial time series we follow Lo et al. [10] that provided a method rooted in a procedure for smoothing the random time series and the numerical search of tops and bottoms using the signal of the derivatives. As far as we could detect in our literature review, the work by Lo et al. was indeed the first one
in which this smoothing method was used for the analysis of financial series. It has served as inspiration for many authors aiming to implement those rules for practical purposes [12, 13]. The current study, however, is focused rather on the understanding of financial fluctuations and, in this sense, differs from previous works that were based on Ref. [10].

It is worth mentioning that our work is also very related to recent works by Preis et al. [14, 15, 16] that has tried to characterize trend switching processes in financial markets. Using a different approach to determine numerically top and bottom prices (a transaction price is considered to be a top (bottom) price if there is no higher (lower) price in a given sliding interval with previously defined size), they study microtrends between these switching points in financial time series. Furthermore, they pose the issue that the understanding of these “micro crises” may help to understand large financial crises that are difficult to model since they are rare events. Although the focus of our paper is different, the methodology based on Lo et al. [10] used here can be also be applied in their context.

The rest of this work is organized as follows: in Section 2 we explain the main steps to the methodology. The discussion of our results is divided into two sections, where we first present the statistics of TB returns and TB intervals (Section 3) followed by a discussion of memory effects (Section 4). Section 5 closes the paper with our concluding remarks.

2. The determination of tops and bottoms

In order to determine the tops and bottoms, we consider a methodology based on Lo et al. [10], which was introduced for automating technical trading
rules. This method is based on two steps: (1) to smooth the time series; and (2) to determine the tops and bottoms using the smoothed time series. In order to smooth the price time series, one assumes that the price of an asset is given by

\[ p_t = y(t) + \varepsilon(t), \]  

where \( y(t) \) is a nonlinear fixed smooth function that depends on time \( t \) and \( \varepsilon(t) \) is the white noise sequence.

We also assume that the estimator \( \hat{y} \) of \( y \) is given by

\[ \hat{y}(t) \equiv \frac{1}{T} \sum_{s=1}^{T} \omega_s(t) P_s \]  

where the weights \( \omega_s \) are larger for those \( P_s \) in which \( s \) is close to \( t \) and are smaller otherwise. Furthermore, the functional form of the weights defines the size of the neighborhood where the average is evaluated, with a clear tradeoff between large and small sizes. A very large one means that the weighted average is very smooth. On the other hand, a very small neighborhood implies that the weighted average is subject to frequent changes. Therefore, using the Gaussian kernel, such as in Lo et al. [10], we assume that the weights are given by

\[ \omega_{s,h}(t) \equiv K_h(t - s)/g_h(t) \]  

where

\[ g_h(t) \equiv \frac{1}{T} \sum_{u=1}^{T} K_h(t - u) \]
and

\[ K_h(x) = \frac{1}{h\sqrt{2\pi}} e^{-x^2/2h^2} \]  \hspace{1cm} (5)

is the Gaussian kernel and \( h \) is the so-called bandwidth – a smoothing parameter that controls the size of the above-mentioned neighborhood.

Therefore, using the definition of \( \hat{y} \) in Eq. (2) and \( \omega \) in Eq. (3), one gets the Nadaraya-Watson kernel estimator

\[ \hat{y}_h(t) = \frac{1}{T} \sum_{s=1}^{T} \omega_{s,h}(t) P_s = \frac{\sum_{s=1}^{T} K_h(t - s) P_s}{\sum_{u=1}^{T} K_h(t - u)} \]  \hspace{1cm} (6)

Based on this procedure, contingent to the choice of \( h \), one can generate a smoothed price time series. The next step is to use this smoothed time series to find the local extremes. One can identify these local extremes by finding the dates \( \tau \) such that \( \text{Sgn}(\hat{y}'_h(\tau)) = -\text{Sgn}(\hat{y}'_h(\tau + 1)) \), where \( \hat{y}'_h \) is the derivative of \( \hat{y}_h \) with respect to \( \tau \) and \( \text{Sgn}(\cdot) \) is the signal function. A positive (negative) value of \( (\hat{y}'_h(\tau)) \) followed by a negative (positive) value of \( (\hat{y}'_h(\tau + 1)) \) means that we have found a local maximum (minimum) \( \hat{y}_h^M (\hat{y}_h^m) \). The local maximums (minimums) are the tops (bottoms) in this work.

3. Distribution of TB returns and TB intervals

In order to obtain a meaningfully large number of TB events in actual data sets, it is necessary to work either with very long inter-day series or with high frequency intra-day records. Therefore, we use our approach to analyze, in first place, the inter day Dow Jones Industrial Average index (DJIA) in
the 1928–2010 interval. This choice is mainly based on the fact that this is a well known data basis, which has been used as benchmark for developing and validating many other methodologies. Intra-day records are subject to daily modulation as the accumulation of new information affects the operations during the first hour after the market opening, while the necessity if concluding foreseen options usually increases the number of operations just before market closure. This requires filtering operations to get rid of such deterministic periodic influence [8]. While we concentrate our discussion on the results for the DJIA set, it worths mentioning that our method has also been applied, with equal success, to the New York Stock Exchange Energy Index (NEI) intra-day data set. Therefore, we also provide a small sample of these results to further attest the reliability of the developed framework.

In Fig. 1, we show the DJIA evolution of whole series (a). Fig. 1b illustrates the time evolution of the logarithm of DJIA (ln(DJIA)). Due to the large time interval spanned by the series and the fact that the DJIA has been roughly increased by a factor 10, the distribution \( p(x) \) of TB event magnitude \( x \) becomes more accurate if we take into account the relative magnitude of a TB event at the time it occurred, what can be obtained by working with the ln(DJIA) rather than the DJIA series. In the inset we show the comparison between the raw and two smoothed data sets obtained according to the procedure described in the previous section. As already commented, more (less) smooth data are obtained larger (smaller) values of \( h \).

Therefore, we consider herein two distinct distributions: (a) the distribution of TB log-return \( p_h^R(x) \), where \( x = |\hat{y}_M^h - \hat{y}_m^h| \) and \( \hat{y}_M^h \) and \( \hat{y}_m^h \) are two consecutive extreme values of the smoothed series of the ln(DJIA); (b) the
Figure 1: Time dependence of DJIA series from 1928 through 2011. Due to the very large time span, effects of both inflation and economic growth are evident. Comparison between panels (a) and (b) makes it evident that the ln(DJIA) series is more suitable for the analysis of TB returns. Panel (c) illustrates the effect of distinct values of $h$ in the smoothing kernel in comparison with the original data (solid line): $h = 3$ ($h = 7$) corresponds to long (short) dashes.
distribution of TB interval $p'_h(x)$, where $x$ is defined as above but $\hat{y}^M_h$ and $\hat{y}^m_h$ refer to the days when the consecutive extreme values of the smoothed series of the ln(DJIA) were observed.

Working with three distinct values $h = 3, 5, \text{and } 7$ we obtain the corresponding total number of events $N^h_e = 1512, 906, \text{and } 664$. Note that such numbers are almost equal to what is obtained if the search for extreme values is performed on the DJIA series (respectively 1508, 902, and 664), hinting at the reliability of the used method for extreme identification. To obtain a more precise information on the functional dependence of $p(x)$ on $x$, we prefer to work with a histogram free method, what amounts to evaluate the integrated distributions $P^h(x > X)$ and $\overline{P}^h(x > X)$ defined as:

\begin{align}
P^h(x > X) &= \int_X^\infty p(x)dx \quad (7) \\
\overline{P}^h(x > X) &= \frac{1}{x} \int_X^\infty p(x)dx \quad (8)
\end{align}

Though both of them distinguish properly exponential from power-law behavior, $P^h(x > X)$ ($\overline{P}^h(x > X)$) is more adequate to provide the correct values of the exponential constant (power law exponent). Intermediate decaying behavior such as stretched exponential behavior \[^8\] can also be inferred from both distributions.

In Fig. 2 we show the behavior of $\overline{P}^h(x > X)$ for the three quoted values of $h$. The obtained curves strongly suggest an asymptotic power law decay. Since it is well known that the generalized $q$-exponential functions defined by

$$\exp_q(x) = [1 + (1 - q)x]_+^{1/(1-q)},$$

(9)
Figure 2: Integrated distribution $P_h(x > X)$ for TB returns obtained from the DJIA series for three distinct values of $h = 3$ (a), 5 (b), and 7 (c). Circles indicate results from the data points. The solid lines correspond to the $q$-Gaussian function $F(x) = \exp_q(-\beta x^2)$. $q = 1.8$ for all values of $h$. $\beta = 540, 350, \text{ and } 250$ for, respectively, $h = 3, 5, \text{ and } 7$.

with $|p|_+ := \max\{p, 0\}$, have been successfully used to deal with interesting issues in economics [17, 18, 19, 20, 21, 22, 23, 24], we looked for best fits to
the data points in terms of these functions. As clearly evidenced in Fig. 2, it turns out that the curves for the $q$-Gaussian $\exp_q(-\beta x^2)$ accurately explain all the points of the evaluated distribution. The value $q = 1.8$ (independently obtained from the three curves) assigns a power law decay for the distribution tail described by the exponent 2.5. The value of $\beta$, which is related to the inverse time scale between TB events, increases monotonically as $h$ decreases.

![Figure 3: Integrated distribution $P^I_h(x > X)$ for TB intervals obtained from the DJIA series for three distinct values of $h = 3$ (circles), 5 (squares), and 7 (triangles). Unlike for the data in Fig. 2, the data cannot be well adjusted by $q$-Gaussians. The log-linear graph hints that the tail follow an exponential decrease.](image)

On the other hand, the distribution of TB intervals has a completely distinct behavior, as shown by the curves for $P^I_h(x > X)$ in Fig. 3. They
suggest an exponential decay for all values of \( h \), with exception of the very short time scales where a plateau similar to those in Fig. 2 is formed.

Fig. 4 illustrates that a similar behavior is observed for \( P_h^R(x > X) \) when we analyze high frequency NEI intra-day data. For the current analysis, we sampled the series at 5 minute interval and followed the same procedure used before [8] to eliminate the daily trend. We have found a smaller value \( q = 1.5 \) which, as in the DJIA results, remain constant for all values of \( \beta \). This value corresponds to a steeper asymptotic power law decay with exponent 4. The found values of \( \beta \) are much larger, indicating that smaller number of 5-minute intervals between TB events. The shown similarity between results for intra-day NEI series and inter day DJIA series is not restricted to the results shown in Figs. 2 and 4. Indeed, NEI series lead to the same exponential dependence for the distribution of TB intervals as one that displayed in Fig. 3 for DJIA series (not shown).

4. Memory effects in TB returns and TB intervals

It is important to look for evidence of predictability among TB events in the analyzed data. A positive answer to this question means that one can learn from TB past history and make decisions to protect and increase profits in his/her investments. To address this question we undertook two distinct procedures: in the first one we evaluated the usual time correlation function of the fluctuations about the mean value \( \bar{x} \):

\[
C(\tau) = \frac{1}{T} \frac{1}{A} \sum_{e=1}^{T} (x(e) - \bar{x})(x(e + \tau) - \bar{x}),
\]
Figure 4: Integrated distribution $P_h(x > X)$ for TB returns obtained from the intra-day NEI series sampled at 5 minute intervals for three distinct values of $h = 3$ (a), $5$ (b), and $7$ (c). Circles indicate results from the data points. As in Fig. 2, the solid lines correspond to the $q$-Gaussian, but now $q = 1.5$ for all values of $h$. The corresponding values of $\beta$ are 5000, 3000, and 2500 for, respectively, $h = 3, 5,$ and $7$. 
where $A = \sum_{e=1}^{T} (x(e) - \bar{x})^2$. Note that we used the argument $e$ (for event) to stress the fact that events occur at different time intervals, and time does not seem to be, at the first place, the proper variable to measure the dependence of the correlations.

Fig. 5 shows the behavior of the correlation function for the TB interval and TB log-return series, respectively $C^I(\tau)$ and $C^R(\tau)$. As observed in the results of the previous section, the behavior is rather robust with respect to changes in the values of $h$. They indicate complete absence of correlation for the TB intervals, and a noticeable correlation of TB log-returns that lasts for some 150 events. It is interesting to observe a small peak around the value 130, recovering from a minimum observed at 80 events. The presence (absence) of correlation are in agreement with the observed distinct behavior of the event distribution shown in the previous section. Similar features have been observed also for the NEI time series, as illustrated in Fig. 5c for the TB log-returns. Much as illustrated in Fig. 5a, the NEI $C^I(\tau)$ function displays no correlation.

In the second approach, we investigated the distribution of time intervals between TB return events of magnitude larger than a given threshold $M_g$. Previous analyses of switching events [8, 14], where the definition of such events differs from the one adopted herein, revealed that the distribution of return times decays slower than exponentially. Here we have evaluated such distribution as function of time intervals ($TI$) as a function of the number of events ($NE$) between TB events larger than $M_g$. If we set $M_g = 0$, the distribution $P^I_{h1}(x > X)$ becomes identical to that of $P^I_{h}(x > X)$ discussed in the previous section. When we increase the value of $M_g$, it is expected.
Figure 5: Time dependence of the correlation function $C^I(\tau)$ for TB intervals (a) and $C^R(\tau)$ for TB returns (b and c). In (a), $C^I(\tau)$ for DJIA data shows to pure noise behavior, in agreement with the former findings in Figs. 3. Results for NEI series (not show) are alike. Correlated patterns for $C^R(\tau)$ are observed when DJIA (b) and NEI (c) data are taken into account. The weaker correlation in (c) agrees with the steeper power law tail shown in Fig. 4 as compared to that in Fig. 2.
that the distributions will differ more and more from $p_h^I(x)$.

Figure 6: Integrated distributions $P_{h}^{TI}(x>X)$ and $P_{h}^{NE}(x>X)$ for return times of TB ln(DJIA) events as a function of actual time interval (a) and number of events (b). In (a), the curves have a tendency to combine an initially flat behavior followed by fat tails that can be described by functions $A \exp^{-\beta x^\gamma}$. In (b), curves decay much faster and cannot be accurately described either or $q$-exponential functions nor by usual exponentials.
In Fig. 6 we show the behavior of the corresponding integrated distributions $P_{h}^{TI}(x > X)$ and $P_{h}^{NE}(x > X)$. Holding $h = 3$, we note that when $M_{g}$ is increased, the departure of $P_{h}^{TI}(x > X)$ from the distribution $P_{h}(x > X)$ ($M_{g} = 0$) occurs in a progressive way. When $M_{g}$ is restricted to the interval $[0.02, 0.10]$, the obtained points are amenable to be adjusted by a the function $A \exp_{q}^{-\beta x^\gamma}$, which leads to a flat behavior when $z$ is small and a tail $\sim x^{\gamma/(1-q)}$, a qualitatively similar behavior to that obtained in the data of the previous section. The adjusted values for $q$ and $\beta$ respectively change, in a monotonic way, from $1.2$ and $0.015$ to $1.8$ and $0.002$ when $M_{g}$ goes from the lower to the upper limit of the quoted interval. On the other hand, the value $\gamma \simeq 1.52$ is subject to only small fluctuations in the same interval of $M_{g}$ values. Regarding the distribution $P_{h}^{NE}(x > X)$, let us remark that it consists of one single point when $M_{g} = 0$, since all TB events are separated by just one event. When $M_{g}$ is increased, the distribution starts to be built. However, $NE$ starts to extend itself over larger values only when $M_{g}$ is close to $0.10$. At this limit, the distribution does not seem to follow an exponential decay, although the precise dependence could not be identified in terms of $q$-exponentials. The results for NEI series are quite qualitatively similar to those shown in Fig. 6, although the values of adjusted parameters depend on the series.

Finally note that, as the value of $M_{g}$ increases, the number of TB events larger than $M_{g}$ decreases. Therefore, the number of points in the distributions shown in Fig. 6 decreases monotonically from $1511$ ($M_{g} = 0$) to $194$ ($M_{g} = 0.10$). For this reason we do not increase the value of $M_{g}$ beyond this limit.
5. Final remarks

In this work we presented a detailed characterization of switching points on financial series based on the identification of top and bottom events of a smoothed curve obtained from the original data. We used the Nadaraya-Watson smoothing kernel introduced by Lo et al. [10], and focused our analyses on the behavior of TB log-returns and TB intervals, time correlations, and return times. We reported results obtained by the developed method in the analysis of the very large inter-day DJIA series in the period 1928-2010. We also tested the method for the NEI intra-day data in the period 2010-2011, and have explicitly shown, for two different measures that the method leads to qualitative similar results, although the quantitative values of the distributions do depend on the data.

We have first shown TB log-returns and TB intervals obey different behavior differently. The probability distribution of TB log-returns have a power law tail, while that of TB intervals follow an exponential decay. We have adjusted the distribution of event magnitude with the help of $q$-Gaussian functions. The least squares fit for values of $q$ lead to a constant value $q = 1.8$, independently of the the width $h$. However, the value of $h$ influences the parameter $\beta$, which is related to the inverse of time scale between TB events. These results hint that the method is robust with respect to the width $h$ of the smoothing kernel. The same features have been observed for the NEI data, although the value $q = 1.5$ indicates a steeper decay in the probability distribution. The larger values of $\beta$ indicate that the time scale of the NEI series in the 5 minute sampling interval is comparatively shorter than that of the inter-day DJIA series.
Persistence of distinct types of behavior was found both when we kept intact the time structure of the series and probed it for correlation and return times. TB intervals have shown to be uncorrelated, while TB log-returns are correlated both for DJIA as well as for NEI series.

Finally, the results for the return time do depend on whether we consider the time difference between TB events of a given magnitude \( M_g \), or when we consider the number of smaller TB events between them. In the first case we can adjust integrated distribution as function of time in terms of \( q \)-stretched exponential functions. Now, the value of the exponent \( \gamma \) does not depend on \( M_g \), while \( q \) changes from \( q = 1 \) to \( q = 2 \) as \( M_g \) increases. By way of contrast, the distribution of TB events as function of the number of smaller events does seem to follow such simple dependency.

We would like to comment that a previous analysis of switching points properties, where the set of points were identified by observing the behavior of volatility, indicated that the distribution of event magnitudes could be explained by stretched exponential functions. The different properties found in the current work suggest that distinct definitions of switching events lead to different properties. In turn, this indicates that a close comparison of different definitions urges a clear understanding of this class of events.

Acknowledgements: This work was partially supported by the following Brazilian funding agencies: CAPES, FAPESB (through the program PRONEX), CNPq (through grants and through the Complex System National Institute for Science and Technology program).


References


Highlights

- Top-bottom approach based on a Gaussian kernel to characterize switching points.
- Characterizes the magnitude of the difference between successive extreme prices.
- Provides statistics of waiting time, i.e., time lag between events.
- Considers also return time, time lag between events with a minimal magnitude.
- Distribution probabilities have been accurately adjusted by q-Gaussians.