Progress Curves and the Prediction of Significant Market Events

Sofia Apreleva[1,*], Neil Johnson[2], and Tsai-Ching Lu[1]

[1] HRL Laboratories, LLC, Malibu, CA
[2] Department of Physics, University of Miami, Miami, FL
svapreleva@hrl.com

Abstract. Progress curves have been used to model the evolution of a wide range of human activities -- from manufacturing to cancer surgery. In each case, the time to complete a given challenging task is found to decrease with successive repetitions, and follows an approximate power law. Recently, it was also employed in connection with the prediction of the escalation of fatal attacks by insurgent groups, with the insurgency "progressing" by continually adapting, while the opposing force tried to counter-adapt. In the present work, we provide the first application of progress curves to financial market events, in order to gain insight into the dynamics underlying significant changes in economic markets, such as stock indices and the currency exchange rate and also examine their use for eventual prediction of such extreme market events.

Keywords: Progress Curve fitting, stock indexes, currency exchange rates, prediction.

1 Introduction

Progress curves have been used to model the evolution of a wide range of human activities -- from manufacturing to cancer surgery. In each case, the time to complete a given challenging task is found to decrease with successive repetitions, and follows an approximate power law. Recently, it was also employed in connection with the prediction of the escalation of fatal attacks by insurgent groups, with the insurgency 'progressing' by continually adapting, while the opposing force tried to counter-adapt. In the present work, we provide the first application or progress curves to the temporal evolution of financial market events, in order to gain insight into the dynamics underlying significant changes in economic indicators, such as stock indices and the currency exchange rate – and also examine their use for eventual prediction of such extreme market events.

Our use of a progress curve function to analyze how specific features develop within a financial time series, is reasonable given that daily changes of stock markets are driven by news from various domains of human activity. News about a rise in the unemployment rate or inflation leads to a drop in stock market prices reflecting the state of the economy. News about civil unrest, riots or insurgencies are also likely to have a negative impact with market reaction expressing fear of impending instability.

K. Glass et al. (Eds.): COMPLEX 2012, LNICST 126, pp. 11–28, 2013.

One example of the influence of the political situation on stock markets is given in Figure 1. Here we show daily changes (%) of IBVC stock index from the Caracas Stock Exchange (Venezuela) in 2012. The stock index rose 99% in the period from January to April, with the main driver for this growth being the absence of reliable information about President Hugo Chavez's health.

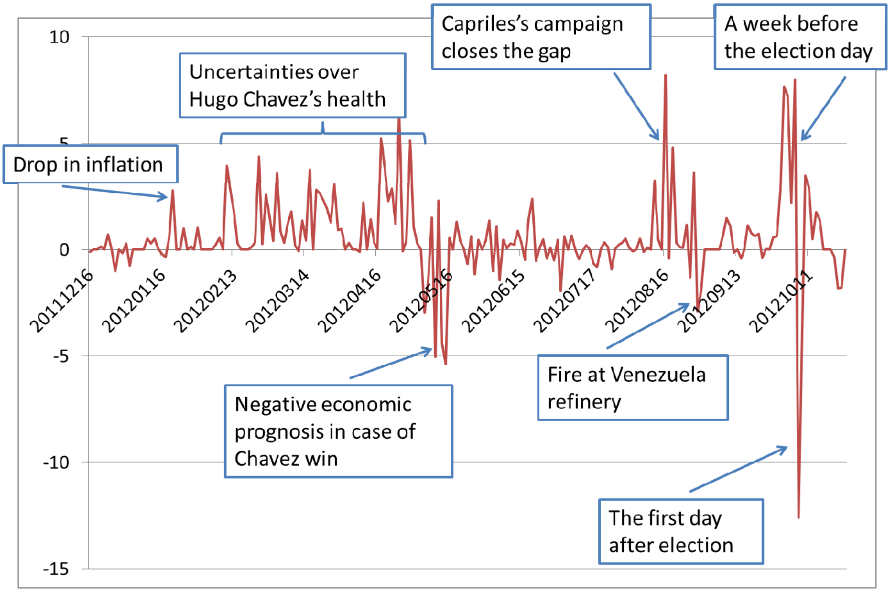

Fig. 1. Daily percent changes in Caracas stock index IBVC for 2012. Text boxes show events related to significant changes.

The arrival of news can stimulate a range of responses among market participants. After they have each weighed the news as being more or less significant for any particular financial instrument such as a stock, their aggregate action then gets recorded as a subsequent market movement. This suggests that the process generating market movements can be viewed as a series of 'blocks', where each block contains a sequence of events related to a particular story. As we can see in Figure 1, the current year 2012 is an election year in Venezuela and the majority of significant events are related to it. It is therefore natural to divide the entire time span into distinct periods with a progress-curve escalation (or de-escalation) dynamics within each. In this paper we show that it is indeed possible to divide the time span into intervals employing the notion of stationary/nonstationary processes. This approach not only avoids assuming that the market obeys stationary dynamics, it represents a conscious effort to capture and quantify the short-term ebbs and flows of trading behavior at the collective level.

We model individual financial time series as a sequence of periods (epochs) with the goal of capturing the successive buildup of behaviors prior to significant market changes. Each of these periods is a combination of escalation processes with switching processes. We then obtain time-series of the Z-score which shows standardized

changes in the original financial time series. These Z-scores are a convenient tool for comparing the behavior of markets in different countries. We focus on Z-score values calculated over certain periods of the original time series (e.g. 30- and 90- day periods). This enables us to partition the Z-score time-series into periods where a stable region is followed by an escalation, before calming down in a final stage. Using progress curves to model the escalation within these periods, we find that our method can predict significant events defined as absolute Z-score value > = 4 for daily time differences with 30 day period, and Z-score value > = 3 with a 90 day period.
In the paper we present explicitly the three steps which comprise the approach:

1. A *time series partitioning* method that derives the periods within which there are distinctive escalation processes. We call these derived periods E-periods.

2. An *escalation parameter fitting method* that trains on historical data within each E-period, in order to derive escalation parameters for the progress curve model;

3. A *progress curve prediction* method that predicts significant events (e.g. absolute Z-score value > = 4 daily time differences with 30 day period, and Z-score value > = 3 with 90 day period) based on parameters obtained at the training stage.

All current methods for modeling and forecasting financial time series can be divided into four categories: 1). statistical modeling of time series; 2). machine learning methods; 3) analysis of influence of social behavior on economic markets (behavioral finance); 4) empirical findings of patterns in time series behavior (most of them obey exponential laws). The most popular data for financial prediction are time series of returns, stock indexes (closing/opening prices), volumes of transactions, interest rates. Here we briefly present the most prominent papers discussing these topics.

Classical econometric approaches for financial markets study and prediction rely on statistical models of time series to reveal the trends, i.e. the connections between consecutive time points. This relation serves as a basis for the prediction of time series values during the next time period. Auto-regression models, hidden Markov models, random walk theory, efficient market hypothesis (EMH) and other models have found large application in the world of finance [4, 9, 10, and 12]. For example, Wang, Wang, Zhang and Guo in Ref. [9] employed a hybrid approach combining three methods for modeling and forecasting the stock market price index. Also used are the exponential smoothing model (ESM), autoregressive integrated moving average model (ARIMA), and the back propagation neural network (BPNN). The weight of the proposed hybrid model (PHM) is determined by a genetic algorithm (GA). The closing of the Shenzhen Integrated Index (SZII) and the opening of the Dow Jones Industrial Average Index (DJIAI) are used as illustrative examples to evaluate the performances of the PHM and to compare with traditional methods including ESM, ARIMA, BPNN, the equal weight hybrid model (EWH), and the random walk model (RWM).

A self-excited multifractal statistical model describing changes of particular time series rather than the time series themselves, was proposed in [14]. There the authors defined the model such that the amplitudes of the increments of the process are expressed as exponentials of a long memory of past increments. The principal feature of the model lies in the self-excitation mechanism combined with exponential nonlinearity, i.e. the explicit dependence of future values of the process on past ones. Distributions of daily changes of stock markets share the same features as distributions of values of the Z score: turbulent flows, seismicity or financial markets, multifractality, heavy tailed probability density functions.

Machine learning models, on the other hand, gained their popularity by incorporating patterns and features obtained from historical data [2, 5]. The authors of Ref. [2] compared three machine learning techniques for forecasting - multilayer perceptron, support vector machine, and hierarchical model. The hierarchical model is made up of a self-organizing map and a support vector machine, with the latter on top of the former. The models are trained and assessed on a time series of a Brazilian stock market fund. The results from the experiments show that the performance of the hierarchical model is better than that of the support vector machine, and much better than that of the multilayer perceptron.

Around 1970, behavioral finance developed into a mature science, focusing the explanation of financial time series variations on the collective behavior of individuals involved into the market [3,7,11]. Recently this vision of how society influences economic indicators has changed to incorporate a broad range of social, political and demographic processes [1,8]. Most of these works has an empirical character in order to overcome the limitations of existing classical approaches.

The work of Bollen, Mao and Zeng [1] explores the influence of society mood states on economic markets, in contrast to behavioral finance which focuses on the collective psychology of individuals involved in sale processes. In particular, Bollen et al. investigate whether measurements of collective mood states derived from large-scale Twitter feeds are correlated to the value of the Dow Jones Industrial Average (DJIA) over time. The text content of daily Twitter feeds is analyzed. A Granger causality analysis and a Self-Organizing Fuzzy Neural Network are used to investigate the hypothesis that public mood states are predictive of changes in DJIA closing values. Their results indicate that the accuracy of DJIA predictions can be significantly improved by the inclusion of specific public mood dimensions, but not others.

Zantedeschi, Damien, and Polson [13] employed dynamic partition models to predict movements in the term structure of interest rates. This allowed the authors to investigate large historic cycles in interest rates and to offer policy makers guidance regarding future expectations on their evolution. They used particle learning to learn about the unobserved state variables in a new class of dynamic product partition models that relate macro-variables to term structures. The empirical results, using data from 1970 to 2000, clearly identify some of the key shocks to the economy, such as recessions. Time series of Bayes factors serve as a leading indicator of economic activity, validated via a Granger causality test.

Polson and Scott [10] propose a model of financial contagion that accounts for explosive, mutually exciting shocks to market volatility. The authors fit the model using country-level data during the European sovereign debt crisis, which has its roots in the period 2008-2010 but was continuing to affect global markets as of October 2011.

Analysis presented in the paper shows that existing volatility models are unable to explain two key stylized features of global markets during presumptive contagion periods: shocks to aggregate market volatility can be sudden and explosive, and they are associated with specific directional biases in the cross-section of country-level returns. Their proposed model rectified this deficit by assuming that the random shocks to volatility are heavy-tailed and correlated cross-sectionally, both with each other and with returns.

The authors of Ref. [8] present a novel approach resulting from studying patterns in transaction volumes, in which fluctuations are characterized by abrupt switching creating upward and downward trends. They have found scale-free behavior of the transaction volume after each switching; the universality of results has been tested by performing a parallel analysis of fluctuations in time intervals between transactions. The authors believe that their findings can be interpreted as being consistent with the time-dependent collective behavior of financial market participants. Taking into account that fluctuations in financial markets can vary from hundreds of days to a few minutes, the authors raise the question as to whether these ubiquitous switching processes have quantifiable features independent of the time horizon studied. Moreover they suggest that the well-known catastrophic bubbles that occur on large time scales—such as the most recent financial crisis—may not be outliers but single dramatic representatives caused by the formation of increasing and decreasing trends on time scales varying over nine orders of magnitude, from very large down to very small.

In contrast to these previous works, the novelty of the approach presented in this paper lies in the unique characterization of system behaviors into escalation and de-escalation periods, and the detection of their switching points. We demonstrate the results of fitting a progress curve to the time series of stock indexes and currency exchange rates in chosen Latin America countries, and make an estimate of the prediction of significant events using this model. The paper presents preliminary results which suggest an extension of the approach in order to take into account the current situation in a country and incorporate filtering/weighting of different events included in the escalation process.

2 Methods and Results

2.1 Overview of the Approach

Figure 2 provides a systematic view of our progress curve prediction system. The time series are partitioned into periods of escalation/de-escalation (E-periods) in order to define time frames for PCM fitting: a stable region is followed by an escalation, before calming down in a final stage. Within each period, sequences of inter-event time intervals and intensities are constructed and the regression parameters Θ_h are found. Prediction of future significant events for the current period can be performed based on regression coefficients obtained from the historical data, or directly from the current E-period if the number of inter-event sequences is >=4.

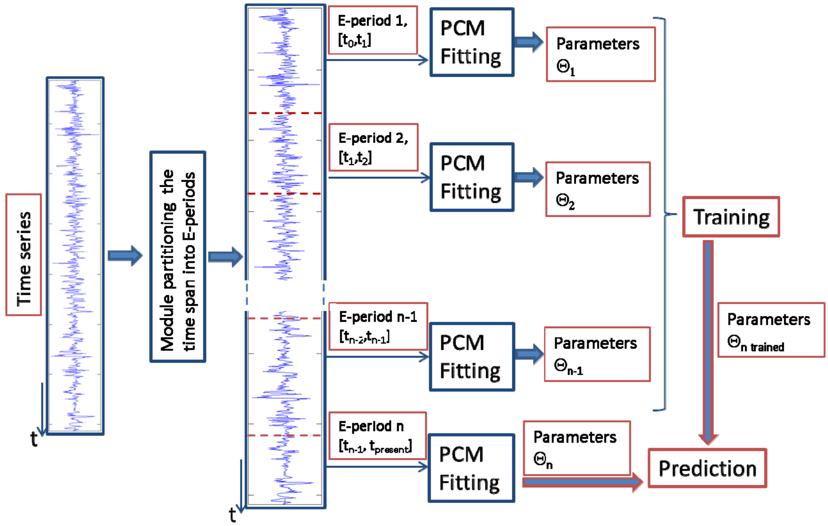

Fig. 2. Schematic overview of significant event prediction using progress curve model

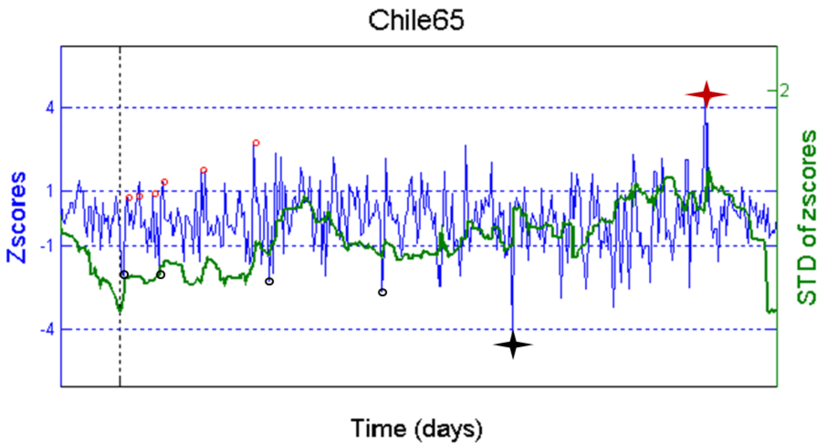

Fig. 3. An event with a Z-score=4.2 for Chile65 stock index, can be predicted using the progress curve formula approximately 220 days in advance. A negative event for the same escalation period with a sigma value of -4.3, can be predicted approximately 65 days in advance. Dashed line shows the beginning of new E- period. Red and black circles mark points forming time series corresponding to positive and negative events**.**

To define distinct periods of Z-score time series, we use the standard deviation (STD) of the Z-score as the measure of market stability. We observe (see Figure 4 and Section 2.2) that a roughly periodical structure emerges when the STD of the Z-score is calculated over different time intervals: the longer the interval, the more evident is the separation into periods of low and high rate-of-change in the Z-score values. We build two time series τ_n and ν_n of points with increasing absolute sigma values >= 1 for inter-event time and inter-event intensity, and then fit progress curve models on a log-log scale. We consider positive and negative significant events separately by fitting separate escalation parameters. We have found that dependences of the regression coefficient on intercepts for each market indicator obey a simple linear relationship.

To make a prediction as to when a significant event with high sigma value will happen in the future, we use simple extrapolation and regression parameters obtained from historical data. The algorithm described above was tested on 16 market indicators (9 stock indices and 7 exchange rates). Below in Figure 3, we give an example of the predictions made for two significant events for Chile65 stock index marked on the figure with a red star (Z-score=4.2) and a black mark (Z-score = -4.3).

2.2 Partition of Z-Score Time Series into E-Periods

In order to characterize significant changes in economic indicators, the raw time series has been transformed to Z-score time series, with time windows of 30 and 90 days. In Figure 4 (top panel) we show an example of Z-score time series for BVPSBVPS index for the time period of July 2007 to July of 2012. It can be seen that the series can be viewed as a sequence of 'waves of burstiness' with a large amplitude of oscillation. We aim to separate these 'waves' and consider them as separate objects of interest representing escalation process followed by the period of calming down. It is an open question if each of these burst periods has to be considered separately, or sometimes their sequence with successfully magnifying amplitude reflects the building up of one large escalation process with unique underlying dynamics. We can assume that each E-period is manifested by a significant event with absolute value of Z-score >= 4 calculated over 30-days' time window and |Z – score| >= 3 of 90-days' time window. This definition implies the unification of several (two or three) burst periods into one E-period.

The following steps are performed to define E-periods:

1. *Derive Z-score time series* from the daily difference of original time series (e.g. daily closing price of BVPSBVPS) with chosen moving time windows (30 or 90 days).
2. *Derive the standard deviation time series* of Z-score time series (called SD time series) with respect to different time windows (e.g., 10, 30, 90 day windows).
3. *Apply standard min/max identification* algorithms to identify local minima and maxima in the identified SD time series from Step 2.
4. *Output E-periods as the duration* of the sequences of local minima and maxima identified in Step 3.

In Figure 4, we show the SD time series (the bottom three panels – 10, 30, 90 days) for the Z-score time series (the top panel) of BVPSBVPS. Step 3 results in the identification of local minima and maxima (dashed lines) that lead to partition and the output of corresponding E-periods.

Intuitively, the daily difference of closing prices represents the first derivative of the original time series. The Z-score time series of this first derivative represents the second derivative - volatility. The SD time series, the standard derivation of Z-score time series, represents the third derivative – momentum of volatility. The identification of local minima and maxima of SD time series naturally give rise to the identification of switching of momentum of volatility – E-periods. This novel partition process enables us to identify periods where the escalation or de-escalation process operates within the identified boundary.

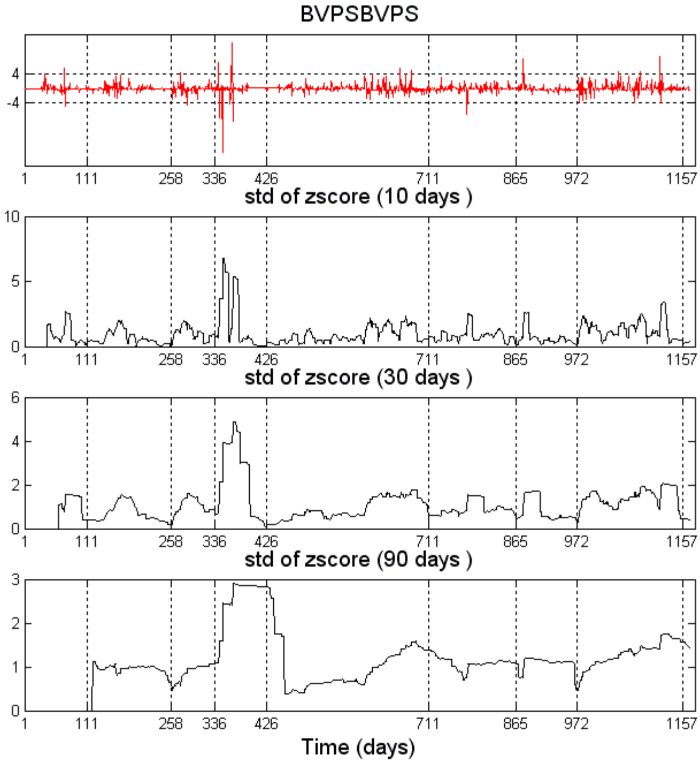

Fig. 4. Illustration of E-periods identification: The top panel shows the BVPSBVPS z-score of 30 day time window. The following three panels show standard deviation of z-score with different time windows: 10, 30, and 90 days. The dashed line partitioned time series into different (de)escalation periods based on minima and maxima of standard deviation of z-score (30 days).

2.3 Progress Curve Modeling

Following Johnson's Progress Curve modeling approach for escalation process [6], we use the following formula:

$$T_n = T_1 n^{-b} \tag{1}$$

where T_n is a time interval between *(n-1)-th* and *n-th* days, *n* is the number of days and *b* is an escalation rate. The challenge in Progress Curve modeling is that behavior process may switch from escalation to de-escalation with different scales and intensity. In the previous section, we presented a novel approach to identify escalation periods, inter-event time and intensity. In this section, we show the steps for escalation parameter fitting modules, which are described as follows:

1. *Extract inter-event time and intensity for escalation trends.* For each period, we identify 'dark' events with absolute vales >= 1 separately for positive and negative sequences. For each 4-sigma (-4-sigma) event, we identify prior 'dark' events to build up the escalation (de-escalation) event trends and then derive inter-event time and intensity. As a result we work with series of τ_n for inter-event times and ν_n for inter-event intensities. Figure5 shows an example of points picked to identify escalation process within an E-period with a following Progress Curve fit for each of constructed sequences.
2. *Fit Progress Curve models to derive escalation parameters.* Given identified inter-event times and intensities for each period, we fit progressive curve model to identify parameters: slope and intercepts (we use the Progress Curve formula in log-log scale). We then fit regression line on (de)escalation parameters from all periods. This gives us a basis for identifying escalation parameter (slop) given intern-event time and intensity (intercepts). We perform the same steps for inter-event intensity. Figure 6 demonstrates the general scheme of algorithm.

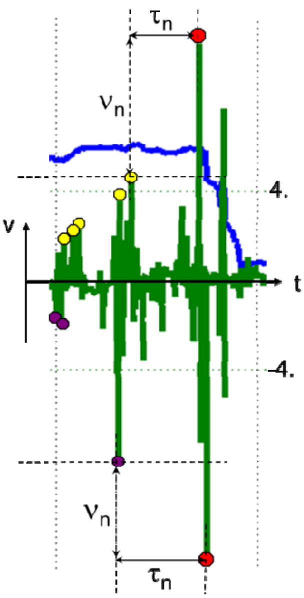

Fig. 5. Illustration of building of inter-event time series τ_n and ν_n. Red circles refer to the significant events, yellow circle - to 'dark' positive events, purple circles denote 'dark' negative eve.

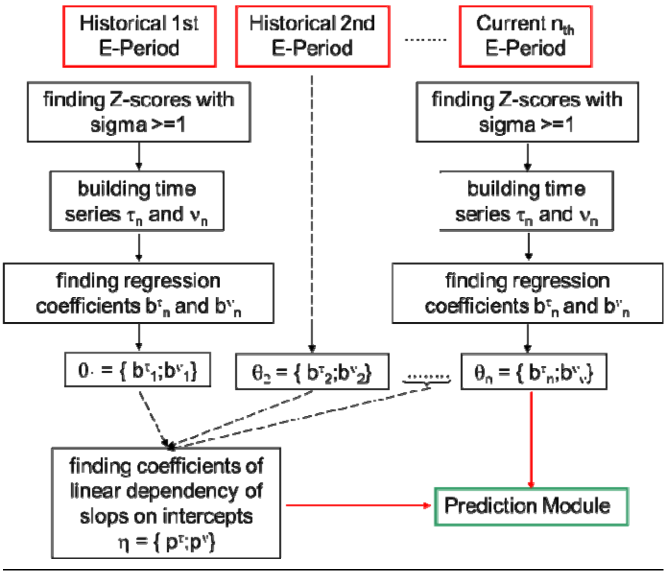

Fig. 6. Scheme of the fitting with PCM. Parameters used in the prediction are obtained from current E-period (θ_n = { $b^{\tau}{}_n$; $b^{\nu}{}_n$} or from the historical data sets (η= {p^{τ} ; p^{ν} }). Fitting performed separately for time series with positive and negative values.

To illustrate the working of the escalation parameter fitting module, we show an example for IGBVL index (July 2007 – July 2012). Results of PCM fitting module of Z-score time series calculated over 90 days period for IGBVL index starting from July 2007 to July 2012 are presented in Figure 7b., which includes pooled together regression coefficients in the form of dependency of slope on intercept, and distribution of correlation coefficients calculated per each E-period.

One can observe that intercepts and slopes obtained from all E-periods obey a simple linear relationship with a reasonably good correlation (see Figure 7). This probably can be explained by the limited range of Z-score values and low frequency of the events to happen. It is known fact that time series of differences of stock market indexes belong to the class of heavy tailed distribution (Figure 7a) with rare occurrences of events with significantly large values. This leads to the limited range of numbers drawn to construct time series building the escalation/excitation process, which leads as a result to the limited range of regression coefficients. We have to note here also that variance of the intercepts is not independent from the one of the slopes: they are related through the variance of the data being subject to the regression.

Once parameters θ_n = { $b^{\tau}{}_n$; $b^{\nu}{}_n$} per each E-period are calculated and parameters η= {p^{τ} ; p^{ν} } characterizing pooled set of regression coefficients is obtained, we can proceed to the module predicting the significant event for the current E-period. Two sets of parameters serve as an input for two different paths used for prediction. The prediction module is described in more detail in the next section.

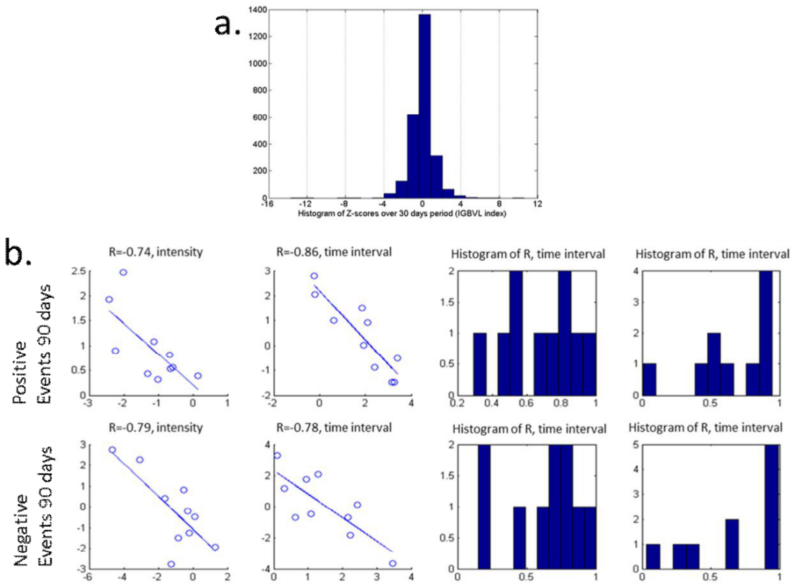

Fig. 7. Figure 7 – Results for PCM fit for IGBVL index: a). basic statistics for inter-event time series τ_n and ν_n. X axis of the first two columns of plots is an intercept of the fit, Y axis is a slop of the fit. Here *R* is a Pearson correlation coefficient; b). typical heavily tailed distribution of Z-score values.

PCM fitting has been performed on 16 financial time series for the same time period (07/2007 – 07/2012) (stock indices and currency exchange rate) given in the Table 1. Table includes the precision criteria as a measure of the accuracy of the prediction, so this metrics will be discussed further in the text.

Table 1. Stock indexes and currency exchange rate used for PCM fitting

Stock Index/ Currency Exchange Rate	**Country**	**Precision**			
		P30	**N30**	**P90**	**N90**
BVPSBVPS	Panama	0.88	0.71	0.42	0.57
CHILE65	Chile	1.00	0.67	0.78	0.73
COLCAP	Colombia	0.75	1.00	0.90	0.89
CRSMBCT	Costa Rica	0.10	0.07	0.10	0.20
IBOV	Brazil	na	0.35	0.42	0.13
IBVC	Venezuela	0.69	0.50	0.64	0.23

Table 1. (*continued*)

IGBVL	Peru	1.00	0.50	0.85	0.91
MERVAL	Argentina	0.60	0.50	0.70	0.75
MEXBOL	Mexico	1.00	0.50	0.67	0.67
USDARS	Argentina	0.67	0.50	0.67	0.75
USDBRL	Brazil	0.50	0.00	0.75	0.00
USDCLP	Chile	0.50	0.00	0.80	0.80
USDCOP	Colombia	1.00	1.00	0.80	0.80
USDCRC	Costa Rica	0.50	0.50	0.80	0.40
USDMXN	Mexico	0.80	0.00	0.75	0.67
USDPEN	Peru	0.57	0.00	0.71	0.50

2.4 Statistical Test of Significance

We performed three significance tests to confirm the validity of fitting financial time series with the progress curve model. We employ time series reshuffling techniques at the level of the entire time span of interest with fixed E-periods (method I); reshuffling of all Z-score values within E-period (method II). Method III works with reshuffled inter-event time and intensities sequences also within particular E-periods.

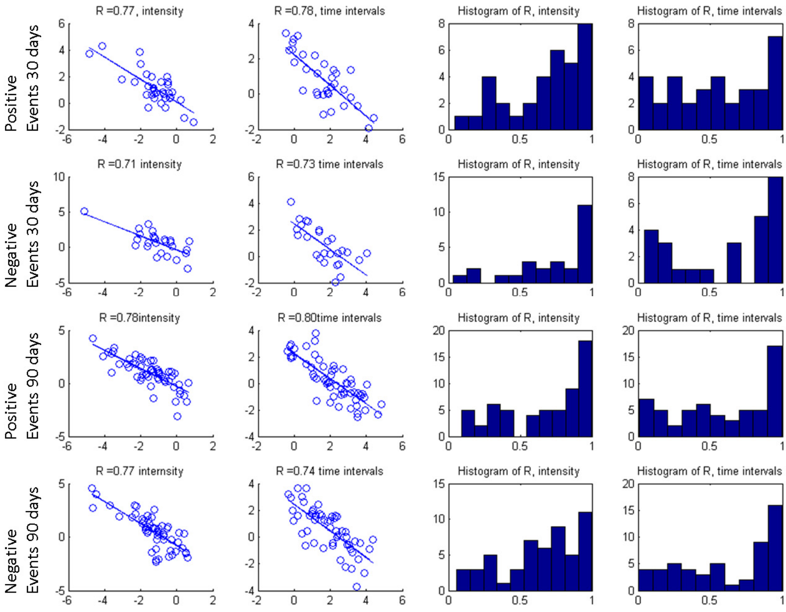

Fig. 8. Results for PCM fit for 16 time series. X axis of the first two columns of plots is an intercept of the fit, Y axis is a slop of the fit.

To carry out the statistical tests, we had to pool parameters $\theta_n = \{b^{\tau}{}_n; b^{\nu}{}_n\}$ per each E-period from all indexes and currency rates to reach a statistical power allowing us to make statistical tests. Pooling together was possible because Z-scores values are standardized metrics as oppose to daily changes of market Z-score, thus they can be used to reveal and investigate behavior and features common for time series representing different indexes. Regression coefficients for all 16 economic indicators and distribution of all corresponding correlation coefficients are shown in the Figure 8. We find here that pooled set of regression parameters reveal properties similar to those observed for individual fits for each indicator. Distributions shown on this figure are used to test for significance and compare against those obtained from randomizing procedures.

Table 2. Results of tests for significance for all types of escalation time series (negative and positive sequences for 30 days Z score time series, negative and positive for 90 days time series). *R* stands for Pearson correlation coefficients, ν_n for inter-event intensities and τ_n for inter-event time intervals.

Pos. 30D

Method	Intercepts ν_n	Slops ν_n	Intercepts τ_n	Slops τ_n	R ν_n	R τ_n
I	0.19481	0.03011	0.00082	0.00193	0.00725	0.00834
II	0.97556	0.38436	0.71327	0.69671	0.49038	0.21443
III	0.00173	0.00011	0.04379	0.04379	0.00008	0.01428

Neg 30D

Method	Intercepts ν_n	Slops ν_n	Intercepts τ_n	Slops τ_n	R ν_n	R τ_n
I	0.02745	0.08955	0.14402	0.72344	0.02222	0.87667
II	0.60838	0.22055	0.8285	0.13333	0.42567	0.01061
III	0.00385	0.01535	0.24464	0.21098	0.00482	0.65643

Pos. 90D

Method	Intercepts ν_n	Slops ν_n	Intercepts τ_n	Slops τ_n	R ν_n	R τ_n
I	0.00001	0.00029	0.22755	0.05098	0.00002	0.67475
II	0.49325	0.57649	0.39936	0.12133	0.40901	0.04833
III	0.00007	0.00001	0.11758	0.04093	0.00001	0.14779

Neg.90D

Method	Intercepts ν_n	Slops ν_n	Intercepts τ_n	Slops τ_n	R ν_n	R τ_n
I	0.00713	0.01248	0.00965	0.10357	0.01189	0.13235
II	0.45183	0.67793	0.13995	0.72016	0.89812	0.79491
III	0.11394	0.02779	0.14126	0.10413	0.14228	0.02081

Table 2 demonstrates p-values for T test (distributions of regression coefficients) and Wicoxon non parametric rank sum test for distributions of correlation coefficients. Tests for regression coefficients were made for intercepts and slopes separately.

Method II (reshuffling of Z score values within E-period) does not show statistical significance. As we can see from the table, the majority of p-values are < 0.05 for method I and method III, which proves that the observed patterns and predictions based on PCM fit does not happen at random.

2.5 Prediction Using PCM

In the previous section we have demonstrated that the progress curve approach can be used to analyze the behavior of financial time series of stock indexes and currency exchange rates: increasing of absolute values of Z-score within one escalation period can be described using progress curve formula to a satisfactory accuracy (see distribution of correlation coefficients for inter-event intensities and time intervals presented in Figure 7b). We also have shown that parameters $\theta_n = \{b^{\tau}_{n}; b^{v}_{n}\}$ per each E-period can be pooled together and dependence of slopes on intercepts $\eta = \{p^{\tau}; p^{v}\}$ can be obtained. To predict a date of the significant event for the current E-period, two approaches are proposed and described in this section. The first set θ_n is used in an extrapolation algorithm and the second set ηenters the second block utilizing the historical data ('trained') linear dependency.

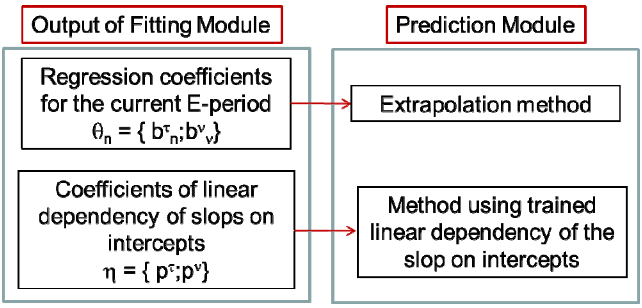

Fig. 9. Scheme of the fitting/prediction module. Parameters used in the prediction are obtained from current E-period ($\theta_n = \{ b^{\tau}_{n}; b^{v}_{n}\}$ or from the historical data sets ($\eta = \{p^{\tau}; p^{v}\}$). Fitting performed separately for time series with positive and negative values.

The simple relationship between inter-event time and intensity provide a tool for prediction of future significant events. As soon as several points building escalation process are observed, we can predict at what moment in the future the significant event exceeding 4 (-4) sigma for 30 days Z – score time series and 3 (-3) for 90 days Z-score time series will happen. To predict this event, a simple formula for integration over inter-event intensities is used:

$$n = \exp\left(\frac{(I - Z_0)(-b_\nu + 1)}{\nu_1} \right) \Big/ (-b_\nu + 1) \qquad (2)$$

where I is a value of significant Z-score; Z_0 is a Z-score value at the time of the first sigma event; b_ν is a slope of the PCM fit for the intensities sequence; ν_1 is an intercept of the fit. Knowing the number of inter-event intervals n and parameters of PCM fit for time interval, one can calculate the final date of the significant event. This formula represents the heart of the extrapolation algorithm. Results of this simple procedure are shown in the Figure 10. Green dots on the picture representing predicted dates are connected with the real dates of events (blue circles).

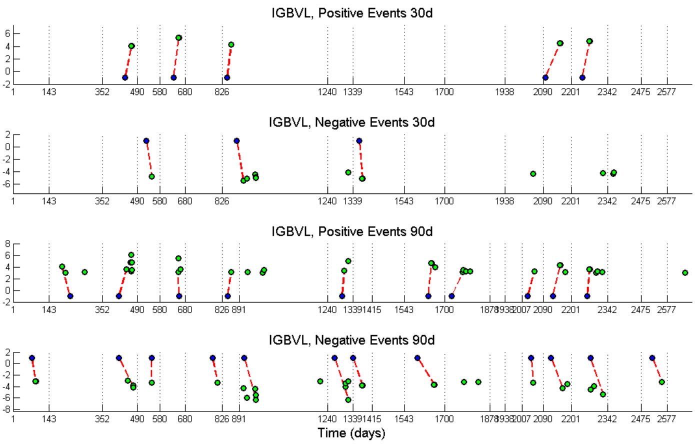

Fig. 10. Results for prediction using extrapolation method: Green circle represent significant events, blue circle – predicted date of event, red line connects predicted and actual events, dashed line separates E-periods

As we mentioned above, the slopes and intercepts of PCM fits for both inter-event time intervals and intensities are in reasonable agreement and regression parameters for this linear relationship can be used for prediction. The advantage of this approach as opposed to an extrapolation procedure, is that one needs only the first time interval between 'dark' events comprising the escalation process. So the final algorithm uses the same formula (2) as the extrapolation method, but regression parameters for both sequences are taken from historical data. To test the prediction using the distribution of regression coefficients obtained from historical data, we have used a longer time span which has been available for stock indexes given in the Table 1 – from the beginning of 2002 to July 2012. This helped us to increase the statistical power since, as we mentioned before, significant events are relatively rare events. Prediction results for IGBVL index using regression coefficients obtained from historical data are displayed on Figure 11.

As can be seen from Figures 10 and 11, the majority of significant events for IGBVL for all 4 types of events can be predicted reasonably well. The following metrics have been calculated to characterize prediction accuracy – precision, D-time and L-time. The two latter metrics serve to estimate the time interval between predicted dates and actual (D-time) and between the time points when prediction is made and the actual time of the event (L-time). Precision is defined as a ratio of events predicted with an absolute value of |D-time|<=50 days. Although at this stage we cannot predict each of multiple significant events within an E-period, this can be done in the future as an extension and refinement of the method. The present approach gives only a rough estimate of the escalation processes. D-times range from 3 to 30 days and L-time usually depends on the duration and constitutes 50 days on average.

In Table 1, the precision metric is displayed. Precision varies from 0 to 1 demonstrating satisfactory performance of the algorithm. It is only for two time series CRSMBCT (Costa Rica) and IBOV (Brazil) that the algorithm shows poor performance and an inability to predict a reasonable amount of events.

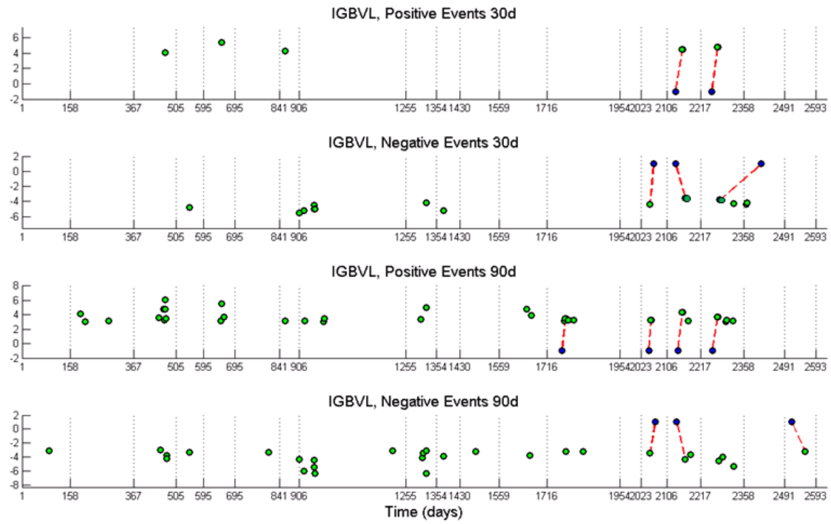

Fig. 11. Results for prediction using training method. Green circles represent significant events, blue circle – predicted date of event, red line connects predicted and actual events, and dashed lines separate E-periods. Two third of all E-periods have been used to obtain distribution of regression coefficients. Prediction is made for the rest of the time span.

3 Conclusions and Future Work

The present work shows the approach to fit financial time series, such as stock indexes and the currency exchange rates, with a progress curve applied within a period with a distinctive build-up of the escalation processes. The idea of this approach has been justified by the modeling of a broad range of human activities using this class of functions, and the main goal was to find progress curve patterns in financial markets.

To enable the fitting we have proposed the algorithm which consists of two steps: 1) partitioning of the entire time series into periods of stationary/nonstationary behavior using standard deviation time series of Z-score values; 2) fitting block which comprises the construction of inter-event time intervals and intensities and fitting itself. Regression coefficients found in the fitting module can be further used for prediction of significant events.

Results of the application of the entire scheme to the financial times series for Latin America countries (9 stock indexes and 6 currency exchange rates) demonstrate the validity of the approach. The results also demonstrate the predictive power of the approach for rough estimation.

Future work can be directed to refine the approach in order to improve the definition of the E-periods, concentrating and incorporating the current situation of the country. This would help to unite or separate several small 'burst' periods depending on their underlying context. Also the studying of news which had the most significant impact, could help to filter non-relevant 'dark' events and improve the predictability of the method. Another possible direction is to build a statistical model which models the escalation processes with a better accuracy.

Acknowledgement. Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI / NBC) contract number D12PC00285. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBE, or the U.S. Government.

References

1. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal of Computational Science 2(1), 1–8 (2011)
2. Carpinteiro, O.A.S., Leite, J.P.R.R., Pinheiro, C.A.M., Lima, I.: Forecasting models for prediction in time series. Artificial Intelligence Review 38(2), 163–171 (2012)
3. Fama, E.F.: Market efficiency, long-term returns, and behavioral finance. Journal of Financial Economics 49(3), 283–306 (1998)
4. Fox, E., Sudderth, E.B., Jordan, M.I., Willsky, A.S.: Bayesian non-parametric inference of switching dynamic linear models. IEEE Transactions on Signal Processing 59(4), 1569–1585 (2011)
5. Huang, W., Nakamori, Y., Wang, S.: Forecasting stock market move-ment direction with support vector machine. Computers and Operations Research 32(10), 2513–2522 (2005)
6. Johnson, N., et al.: Pattern in escalations in insurgent and terrorist activity. Science 333(6038), 81–84 (2011)
7. Lu, B., Song, X., Li, X.: Bayesian analysis of multi-group nonlinear structural equation models with application to behavioral finance. Quantitative Finance 12(3), 477–488 (2012)
8. Preis, T., Schneider, J.J., Stanley, H.E.: Switching processes in financial markets. Proceedings of the National Academy of Sciences of the United States of America 108(19), 7674–7678 (2011)

9. Wang, J., Wang, J., Zhang, Z., Guo, S.: Stock index forecasting based on a hybrid model. Omega 40(6), 758–766 (2012)
10. Polson, N.G., Scott, J.C.: Predictive macro-finance with dynamic partition models (2011), http://faculty.chicagobooth.edu/nicholas.polson/research/papers.html
11. Shiller, R.J.: From efficient markets theory to behavioral finance. Journal of Economic Perspectives 17(1), 83–104 (2003)
12. Watanabe, K., Takayasu, H., Takayasu, M.: A mathematical definition of the financial bubbles and crashes. Physica A: Statistical Mechanics and its Applications 383(1 spec. iss.), 120–124 (2007)
13. Zantedeschi, D., Damien, P., Polson, N.G.: Predictive macro-finance with dynamic partition models. Journal of the American Statistical Association 106(494), 427–439 (2011)
14. Filimonov, V., Sornette, D.: Self-excited multifractal dynamics. EPL (Europhysics Letters) 94(4), 46003 (2011)