Computational Analysis and Stock Price Modelling by Interacting Contact System

GE YANG
Beijing Jiaotong University
School of Science
Shangyuancun 3, Beijing 100044
P.R. CHINA

JUN WANG
Beijing Jiaotong University
School of Science
Shangyuancun 3, Beijing 100044
P.R. CHINA
wangjun@bjtu.edu.cn

Abstract: We perform the computational analysis on the fluctuations of financial stock market. We establish a stock price model based on the stochastic contact system. The contact system is a member of a class of stochastic processes known as interacting particle systems, and it is also a model for epidemic spreading that mimics the interplay of local infections and recovery of individuals. Then, we introduce a new method detecting the duration and intensity relationship in volatility series. At last, we analyze the Zipf behaviors of Shanghai Stock Exchange (SSE) Composite Index, and the simulative data derived from the stock model by comparison in different time scales and different threshold scales.

Key–Words: Computational analysis, stock price model, contact system, computer simulation

1 Introduction

Investigating the volatility behaviors of financial time series has long been a focus of financial research. Financial market is a complex evolving system which has many influence factors and many kinds of uncertainties, its fluctuation usually represents strong nonlinear characteristics [1, 3-7, 9-25]. And it is becoming a key issue to model the dynamics of the forwards prices in the risk management, derivatives pricing, physical assets valuation and forecasting. Any model aiming at understanding price fluctuations needs to define a mechanism for the formation of the price, in an attempt to create the representations of reality. Economic systems such as financial markets are similar to physical systems in that they are comprised of a large number of interacting agents, and these economic agents interact in complicated ways. Recently, some agent based financial models have been established by statistical physics systems [2, 8], which attempt to explain a financial market as an aggregation of single agents, or clusters of agents, with specific roles assigned. For instance, the financial time series models are developed by the percolation and oriented percolation systems, in which the local interaction or influence among traders in a stock market is constructed, and a cluster of percolation is applied to define the cluster of traders sharing the same opinion about the market [13, 14, 21, 22], where the critical phenomenon of percolation is used to illustrate the herd behavior of stock market participants. And an agent price model based on the finite-range contact particle system is proposed to explore the power-law tail distribution, the long range correlation and the q-Gaussian distribution of normalized returns.

The stochastic contact process, one of statistical physics systems [2, 8], is a model for epidemic spreading in a continuous time Markov process. In the present work, the stochastic contact system is applied to model the dynamics of stock prices, in an attempt to study the nonlinear phenomena of volatility duration series of stock returns, where a new analyzing method is introduced to transfer volatility series into volatility duration series, and we apply it to investigate the corresponding duration patterns. For the empirical research of volatility duration series, we investigate the corresponding Zipf behaviors and cross-correlation behaviors for the simulation data of the price model and the real markets data by the statistical analysis. The daily closing prices data of Shanghai Stock Exchange (SSE) Composite Index and Shenzhen Stock Exchange (SZSE) Component Index are selected as the real empirical data.

2 Description of stock price model from contact system

The contact system is often thought of as a crude model for the spread of a disease or a biological population. Healthy individuals become infected at a
where at time Let $\delta < \lambda < \lambda$ is an individual at the point $x$ is healthy and will be infected at a rate equal to $\lambda$ times the number of the infected neighbors. For functions $f$ on $(0, 1)^{Z^d}$ that depend on finitely many coordinates, the generator of the dynamics of contact process is defined by

$$A f(\eta) = \sum_x c(x, \eta)[f(\eta^x) - f(\eta)],$$

where $\eta^x(y) = \eta(y)$ if $y \neq x$, $\eta^x(y) = 1 - \eta(x)$ if $y = x$, for $x, y \in Z^d$. And $c(x, \eta)$ is the transition rates which are given by

$$c(x, \eta) = \begin{cases} 1, & \text{if } \eta(x) = 1, \\ \lambda \sum\{y : |y - x| = 1\} \eta(y), & \text{if } \eta(x) = 0. \end{cases}$$

Let $\eta^A_s$ denote the state at time $s$ with the initial state $\eta^A_0 = A$. Further let $\eta^{(0)}_s(x)$ be the state of $x \in Z^d$ at time $s$ with the initial point $\{0\}$. The most important feature of the contact process is that survival and extinction can be both occur, which of these occurs depends on the value of $\lambda$. There is a critical point $\lambda_c = \inf\{\lambda : P(\eta^{(0)}_s > 0, \forall s \geq 0) > 0\}$, where $|A|$ denote the cardinality of a finite set $A$. If $\lambda < \lambda_c$, the contact process is said to die out or become extinct, that is, $P(\eta^{(0)}_s = 0, \forall s \geq 0) > 0$; otherwise, for $\lambda > \lambda_c$, it is said to survive.

We consider $Z \times R_+$ as the space-time, for each pair $x, y \in Z$ with $|x - y| = 1$, let $\{T_n(x, y) : n \geq 1\}$ be a Poisson process with rate $\lambda$, and let $\{U_n \in N : n \geq 1\}$ be a Poisson process with rate $1$. At times $T_n(x, y)$, we draw an arrow from $x$ to $y$ to indicate that if $x$ is infected then $y$ will become infected (if it is not already). At times $U_n$, we put a $\delta$ at $x$. The effect of a $\delta$ is to recover the individual at $x$ (if one is infected). To construct the process from this “graphical representation”, we imagine fluid entering the bottom side at the points in $\eta_0$ (at time 0) and flowing upward the structure. The $\delta$’s are the dams and the arrows are pipes which allow the fluid to flow in the indicated direction. Let $\tau = \inf\{s \geq 0 : \eta^{(0)}_s = 0\}$. For $\lambda > \lambda_c$, then $\lim_{s \to \infty} \eta^{(0)}_s / s = 2\alpha(\lambda) \rho(\lambda)$, a.s. on $\{\tau = \infty\}$, where $\rho(\lambda) \geq 0$ is a nondecreasing function of $\lambda$, and $\alpha(\lambda) \geq 0$. And if $\lambda < \lambda_c$, for some positive $\rho(\lambda)$, we have $P(\eta^{(0)}_s = 0) \leq e^{-\rho s}$, then the process dies out exponentially fast. Furthermore, for any large positive integer $L \geq 1$, let $T_{\eta^A_s}$ be the truncated contact process defined via the graphical representation, but using only paths with vertical segments corresponding to sites in $(-L, -L)_d$ and infection arrows from $(x, \cdot)$ to $(y, \cdot)$ with $x \in (-L, -L)_d$. Then, for every finite $A$ and constant $C \geq 1$

$$\lim_{s \to \infty} \lim_{L \to \infty} P(\|T_{\eta^A_s} \geq C) = P(\eta^A_s \neq \emptyset, \forall s > 0).$$

More generally, we consider the initial distribution as $\nu(\rho)$, the product measure with density $\rho$, that is, each site is independently occupied with probability $\rho$, and let $\eta^A_0$ to denote the contact model with initial distribution $\nu(\rho)$.

In the present work, we assume that the stock price fluctuation results from the investors’ investment attitudes towards to the stock market, and suppose that the investment attitude is represented by the viruses of the contact model, which accordingly classify the market participants with buying position, selling position and holding position respectively. Consider a model of auctions for a stock in a stock market. Assume that each trader can trade the stock several times at each day $t \in \{1, 2, \cdots, T\}$, but at most one unit number of the stock at each time. Let $T$ be the time length of trading time in each trading day, we denote the stock price at time $s$ in the $t$th trading day by $\bar{P}(s)$ where $s \in [0, l]$. Suppose that this stock consists of $2M + 1(M$ is large enough) investors, who are located in a line $\{-M, \cdots, -1, 0, 1, \cdots, M\} \subset Z$ (similarly for $d$-dimensional lattice $Z^d$). At the beginning of trading in each day, suppose that the investor at the original point receives the investing information. We define a random variable $\xi(\rho)$ for this investor, suppose that this investor takes buying position $\xi = 1$, selling position $\xi = -1$ or neutral position $\xi = 0$ with probability $p_1, p_{-1}$ or $1 - (p_1 + p_{-1})$, respectively. Then this investor sends bullish, bearish or neutral signal to his neighbors according to the contact dynamic system. Investors can affect each other or the information can be spread, which is assumed as the main factor of price fluctuations. For a fixed $s \in [0, l]$, the aggregate excess demand for the asset at time $t$ is defined by

$$\bar{\mathcal{B}}_t(s) = \xi(\rho)|\eta^A_s|/M.$$
where $\alpha > 0$, represents the depth parameter of the market, and $P_0$ is the initial stock price at time 0. The formula of the single-period stock logarithmic returns from $t$ to $t+1$ is given as follows

$$R_t = \ln P_{t+1}(s) - \ln P_t(s).$$

### 3 Zipf behaviors of volatility duration financial series

#### 3.1 Volatility duration financial series

The analysis of volatility behaviors of financial series is an active topic in financial research, for instance, there exists a return interval analysis that it characterizes the occurrence of volatilities in certain range. In the following, a volatility duration analysis is developed to study the financial volatility series (or the absolute return series). We are interested in the duration of the stock volatility consistently above or below a given data point in the volatility series. In this analysis, the threshold is not predetermined but is changing with the volatility intensity along the series. From this method, we make an approach to build a link between volatility intensity and duration, and hope to acquire some insights by this linkage. Actually, this proposed analysis share much similarity with the ideas embodied in intensity-duration-frequency relationship, which is a method widely used in description of data records of precipitation, climate, flood and wave, etc. We think that this approach is useful for stock volatility series analysis.

![Fig. 1. The volatility duration series from SSE composite index.](image)

Considering a volatility duration series $D_t$ ($t = 1, \ldots, T$), which is derived from the corresponding absolute return series $|R_t|$. At each trading day $t$, we intend to investigate the absolute return of the next day $|R_{t+1}|$. We say that the volatility series is locally rising at $t$, when $|R_{t+1}| > |R_t|$, then the time duration of volatility intensity exceeding $|R_t|$ as the volatility duration length $I(t)$ at day $t$, namely, $I(t) = \max\{\tau : |R_{t+\tau}| > |R_t|, \forall \tau \leq t\}$. And we say that the volatility series is locally falling at $t$, when $|R_{t+1}| < |R_t|$, then we have $I(t) = \max\{\tau : |R_{t+\tau}| < |R_t|, \forall \tau \leq t\}$. For the case $|R_{t+1}| = |R_t|$, we let $I(t) = 0$. Also, in order to moderate extreme values in volatility duration series, we take the square root of $I(t)$ as the final value of each data point in volatility duration series and assign each one a positive or negative sign. $D_t = [I(t)]^{\frac{1}{2}}$ if $I(t)$ describes a duration of exceedance, otherwise, $D_t = -[I(t)]^{\frac{1}{2}}$. An entire series of absolute returns can be scanned in this way, and we obtain duration series $D_t$ consequently. The plot of the volatility duration series from SSE composite index is displayed in Fig. 1.

#### 3.2 Introduction of Zipf analysis

Zipf analysis, originally introduced in the context of natural languages to study the statistical occurrences in different languages, has been applied to various types of data in physical and social sciences [12, 21]. Recently Zipf analysis is revealed as as a way for quantifying time series correlations. The technique is based on translating a given time series into a sequence of symbols and counting the frequency of any word, that is, pattern of consecutive symbols. Ranking these words by their frequencies from most common to least common, and plotting the logarithm of frequencies versus the logarithm of rank give us a Zipf plot. In this section, we apply an extended Zipf-type method to investigate the symbolic dynamics of the financial model derived from the contact model with the parameters $\lambda = 1.2$ and $\rho = 0.6$, and comparatively study the corresponding results with that of SSE data. The definition of $\tau$-return of stock prices is given as

$$R_t(\tau) = \ln P_{t+\tau} - \ln P_t, \quad t = 1, \ldots, T - \tau.$$  

For $\tau \in \{1, 5, 20, 60, 250\}$, we call them characteristic time scales, which approximately stand for one transaction day, one transaction week, one transaction month, one transaction quarter and one transaction year respectively in terms of business time units (with weekends and holidays eliminated). From Sect. 3.1, we can obtain the corresponding volatility duration series $D_t(\tau)$ from the absolute return series $|R_t(\tau)|$. Next the $\tau$-duration series is transformed into a new three-alphabeted symbolic sequence $y_t(\tau, \theta)$ as follows

$$y_t(\tau, \theta) = \begin{cases} u, & \text{if } D_t(\tau) \geq \theta \\ s, & \text{if } |D_t(\tau)| < \theta \\ d, & \text{if } D_t(\tau) \leq -\theta \end{cases}$$
where \( u, s \) and \( d \) denote “duration-up”, “duration-stable” and “duration-down” respectively. \( \theta \) is a variation threshold, which can be interpreted as the expected duration for the market, which, in this analysis, is a nonnegative random variable on a probability space (with the probability distribution function \( F_\theta(x) \)). For instance, it can be a uniform distribution on the interval \((0, 2)\).

We introduce the absolute frequency and the relative frequency to analyze the statistical behaviors of \( y_t(\tau, \theta) \) for different values of parameters \( \tau \) and \( \theta \). Let \( n_u(\tau, \theta) \), \( n_s(\tau, \theta) \) and \( n_d(\tau, \theta) \) denote the frequencies of occurrences for “duration-up”, “duration-stable” and “duration-down” respectively. Then the corresponding absolute frequencies of sequence \( y_t(\tau, \theta) \) are given as follows

\[
f_u(\tau, \theta) = \frac{n_u(\tau, \theta)}{T - \tau} \times \frac{1 - F_\theta(x)}{2}
\]
\[
f_d(\tau, \theta) = \frac{n_d(\tau, \theta)}{T - \tau} \times \frac{1 - F_\theta(x)}{2}
\]
\[
f_s(\tau, \theta) = \frac{n_s(\tau, \theta)}{T - \tau} \times F_\theta(x)
\]

and the corresponding relative frequencies are given as

\[
g_u(\tau, \theta) = \frac{n_u(\tau, \theta)}{n_u(\tau, \theta) + n_d(\tau, \theta)} \times (1 - F_\theta(x))
\]
\[
g_d(\tau, \theta) = \frac{n_d(\tau, \theta)}{n_u(\tau, \theta) + n_d(\tau, \theta)} \times (1 - F_\theta(x))
\]

In definitions of the relative frequencies, we neglect the occurrences of duration-stable and use \( g_u(\tau, \theta) \) and \( g_d(\tau, \theta) \) to measure the total occurrences of duration rising and duration falling respectively. In the following, for both the simulation data and the actual data, we consider the statistical properties of absolute frequencies and relative frequencies for different values of two parameters \( \tau \) and \( \theta \).

### 3.3 Zipf analysis for different threshold scales

Fig. 2 shows the frequencies \( f_u(\tau, \theta) \), \( f_s(\tau, \theta) \), \( f_d(\tau, \theta) \) as a function of the time-scale \( \tau \) for four different values of the threshold \( \theta \) for volatility duration series from SSE composite index and the financial model where \( \theta = 2, 4, 6, 8 \). The fluctuations of the absolute frequencies of volatility duration series from SSE composite index are exhibited in Fig. 2(d), (e) and (f), and the corresponding fitting fluctuations of the absolute frequencies of volatility duration series form the price model are displayed in Fig. 2(a), (b) and (c). From these plots we find that, for both the actual data and the simulation data, \( f_u(\tau, \theta) \) decays exponentially to achieve a steady-state value. This means that long-run transactions are less sensitive to transactions costs than short-run transactions. \( f_u(\tau, \theta) \) increases exponentially as \( \tau \) increases till some point then decays linearly as \( \tau \) increases, while \( f_u(\tau, \theta) \) always increases exponentially as \( \tau \) increases. We also note that the curve of \( f_u(\tau, \theta) \) with the smaller threshold scale \( \theta \) lies above the one of \( f_u(\tau, \theta) \) with the larger value \( \theta \). The curves of \( f_d(\tau, \theta) \) exhibit the similar behaviors to those of \( f_u(\tau, \theta) \), whereas the curves of \( f_s(\tau, \theta) \) again show the opposite behaviors.

![Fig. 2](image-url)

**Fig. 2.** (a)(b)(c) are the absolute frequencies for the volatility duration series from simulation data of the financial model. (d)(e)(f) are the absolute frequencies for volatility duration series SSE composite index.

![Fig. 3](image-url)

**Fig. 3.** (a) Relative frequencies of SSE composite (b) Relative frequencies of simulation data with the parameters \( \lambda = 1.2 \) and \( \rho = 0.6 \).

The distributions of the relative frequencies for different expected return durations \( \theta = 2, 4, 6, 8 \) with
various time scales $\tau$ are demonstrated in Figure 3. In Fig. 3, we can find that, for both the actual data and the simulation data, the curves of $g_d(\tau, \theta)$ are mostly larger than 0.5 while the curves of $g_u(\tau, \theta)$ are primarily smaller than 0.5. Furthermore, we note that the curve of $g_d(\tau, \theta)$ with the larger threshold scale $\theta$ lies above the one of $g_u(\tau, \theta)$ with the smaller value $\theta$, whereas the curves of $g_u(\tau, \theta)$ again show the opposite behaviors.

### 3.4 Zipf analysis for different time scales

Table 1 shows the values of inflection points of both absolute frequencies $f_u(\tau, \theta)$, $f_b(\tau, \theta)$, $f_d(\tau, \theta)$ and relative frequencies $g_u(\tau, \theta)$, $g_d(\tau, \theta)$ of volatility duration series are plotted for different time scales $\tau$-duration series, where $\tau = 1, 5, 20, 60, 250$. And the values of inflection points of five frequency functions are divided into two groups for SSE data and the simulation data.

<table>
<thead>
<tr>
<th>Data</th>
<th>$\tau$</th>
<th>$f_u(\tau, \theta)$</th>
<th>$f_b(\tau, \theta)$</th>
<th>$f_d(\tau, \theta)$</th>
<th>$g_u(\tau, \theta)$</th>
<th>$g_d(\tau, \theta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>1</td>
<td>11.4</td>
<td>10.8</td>
<td>11.0</td>
<td>24.8</td>
<td>24.8</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>12.8</td>
<td>12.2</td>
<td>12.4</td>
<td>28.6</td>
<td>28.6</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>15.2</td>
<td>13.6</td>
<td>14.2</td>
<td>32.2</td>
<td>32.2</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>16.8</td>
<td>16.2</td>
<td>16.4</td>
<td>34.2</td>
<td>34.2</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>22.2</td>
<td>18.8</td>
<td>20.2</td>
<td>37.6</td>
<td>37.6</td>
</tr>
<tr>
<td>S.D.</td>
<td>1</td>
<td>11.2</td>
<td>9.2</td>
<td>8.2</td>
<td>25.0</td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>12.4</td>
<td>11.6</td>
<td>11.4</td>
<td>27.4</td>
<td>27.4</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>14.4</td>
<td>13.2</td>
<td>12.8</td>
<td>30.6</td>
<td>30.6</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>17.8</td>
<td>15.8</td>
<td>15.6</td>
<td>32.8</td>
<td>32.8</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>20.2</td>
<td>16.4</td>
<td>17.8</td>
<td>35.6</td>
<td>35.6</td>
</tr>
</tbody>
</table>

1 S.D. means the simulative data.

For each of the functions in Table 1, the value of inflection point becomes larger when the time interval $\tau$ becomes larger. This means that, for any time interval $\tau$, there exists a critical value of $\theta$ for a short risk expected return. If an investor wants to obtain a return that exceed this value, he will face a significantly longer investing risk. If an investor increase his investment time interval, say from 60 days to 120 days, he may enjoy a larger range of shorter risk expected return, that is, he may have higher return without suffering from a higher investing risk.

### 4 Conclusion

In the paper, we introduce a stock price model by applying the contact system to study the behaviors of the fluctuations for the stock markets. We discuss the properties of the model for the intensity $\lambda$ and the radius $\rho$. Then, a new method detecting the duration and intensity relationship in volatility series is introduced. At last, we compare the Zipf behaviors of volatility duration series from Shanghai Stock Exchange (SSE) Composite Index, and the simulative data derived from the stock model in different time scales and different threshold scales. From the discussion, for both actual data and simulation data, we find that for any time interval $\tau$, there exists a critical value of $\theta$ for a short risk expected return.

Acknowledgements: The authors were supported in part by National Natural Science Foundation of China Grant No. 71271026 and Grant No. 10971010. The authors would like to thank the support of Institute of Financial Mathematics and Financial Engineering at Beijing Jiaotong University.

References:


