5
Why Care about a Power Law?

"The reason we care about power law correlations is that we're conditioned to think they're a sign of something interesting and complicated happening."
– Cosma R. Shalizi, *Power-law distributions, 1/f noise, long-memory time series.* http://www.cs.umich.edu/~crshalizi/notebooks/power-laws.html

Social scientists have often remarked that physicists working in social or economic phenomena seem to be obsessed with power laws. Although there are no a priori reasons why, say, log-normal distributions should be any less interesting, it remains a fact that many papers in econophysics are about distributions exhibiting a power-law tail. The reason for this attraction to power laws is that it signals the occurrence of scale-independent behavior, which for physicists has an intimate connection to phase transitions and critical phenomena. In contrast, most other distributions, such as the Gaussian or the Poisson, are characterized by a "typical value" (or "scale") about which all the recorded data points occur. Thus, the absence of such a characteristic scale in a process would indicate that the corresponding variable may vary over a wide range of values, sometimes spanning several orders of magnitude [1]. A large number of examples for power-law distributions occurring in reality have been put forward, such as the size distribution of cities, the intensity of earthquakes and the frequency of word usage in human languages, to name a few. However, none of these are beyond controversy. In fact, critics would say that most of the so-called power laws reported in the scientific literature are an artifact of limited datasets and crude estimation methods [2]. In econophysics, the most reliable examples of possible power-law distributions occur in the financial arena, and in the following sections we shall look at these in detail.

5.1
Power Laws in Finance

Financial markets can be viewed as complex systems with a large number of interacting components that are subject to external influences or information flow. Physicists are being attracted in increasing numbers to the study of financial mar-

Econophysics. Sitabhra Sinha, Arnab Chatterjee, Anirban Chakraborti, and Bikas K. Chakrabarti

ISBN: 978-3-527-40815-3

kets by the prospect of discovering universalities in their statistical properties [3–5]. This has partly been driven by the availability of large amounts of electronically recorded data with very high temporal resolution, making it possible to study various indicators of market activity.

5.1.1
The Return Distribution

One of the principal statistical features characterizing the activity in financial markets is the distribution of fluctuations of individual stock prices and overall market indicators such as the index. To study these fluctuations such that the result is independent of the scale of measurement, we define the logarithmic return for a time-scale Δt as

$$R(t, \Delta t) \equiv \ln P(t + \Delta t) - \ln P(t) , \tag{5.1}$$

where $P(t)$ is the price (or market index) at time t and Δt is the time-scale over which the fluctuation is observed. This definition of price change has the added advantage that to obtain the return at a longer time-scale one simply needs to add together the changes occurring at the shorter time-scale.

Market indices, rather than individual stock prices, have been the focus of most previous studies, as the former are more easily available, and also gives overall information about the market. By contrast, individual stocks are susceptible to sector-specific, as well as, stock-specific influences, and may not be representative of the entire market. These two quantities, in fact, characterize the market from different perspectives, the microscopic description being based on individual stock price movements, while the macroscopic point of view focusses on the the collective market behavior as measured by the market index.

The importance of interactions among stocks, relative to external information, in governing market behavior has emerged only in recent times. The earliest theories of market activity, for example Bachelier's random walk model, assumed that price changes are the result of several independent external shocks, and therefore, predicted the resulting distribution to be Gaussian [6]. As an additive random walk may lead to negative stock prices, a better model would be a multiplicative random walk, where the price changes are measured by logarithmic returns [7]. While the return distribution calculated from empirical data is indeed seen to be Gaussian at long time-scales, at shorter times the data show much larger fluctuations than would be expected from this distribution [8]. Such deviations were also observed in commodity price returns, for example in Mandelbrot's analysis of cotton price, which was found to follow a Levy-stable distribution [9]. However, it contradicted the observation that the distribution converged to a Gaussian at longer time-scales. Later, it was discovered that while the bulk of the return distribution for the S & P 500 index appears to be fit well by a Levy distribution, the asymptotic behavior shows a much faster decay than expected. Hence, a truncated Levy distribution, which has exponentially decaying tails, was proposed as a model for the distribution of returns [10]. Subsequently, it was shown that the tails of the cumulative

return distribution for this index actually follow a power law

$$P_c(r > x) \sim x^{-\alpha} , \tag{5.2}$$

with the exponent $\alpha \approx 3$ (the "inverse cubic law") [11], well outside the stable Levy regime $0 < \alpha < 2$. This is consistent with the fact that at longer time-scales the distribution converges to a Gaussian. This "inverse cubic law" had been reported initially for a small number of stocks from the S & P 100 list [12]. Later, it was established from statistical analysis of stock returns in the German stock exchange [13], as well as for three major US markets, including the New York Stock Exchange (NYSE) [14]. The distribution was shown to be quite robust, retaining the same functional form for time-scales of up to several days. Similar behavior has also been seen in the London Stock Exchange [15]. An identical power-law tail has also been observed for the fluctuation distribution of a number of market indices [16, 17]. This apparent universality of the distribution may indicate that different markets self-organize to an almost identical nonequilibrium steady state.

These observations are somewhat surprising, although not at odds with the "efficient market hypothesis" in economics, which assumes that the movements of financial prices are an immediate and unbiased reflection of incoming news and future earning prospects. To explain these observations various multiagent models of financial markets have been proposed, where the scaling laws seen in empirical data arise from interactions between agents [18]. Other microscopic models, where the agents (i.e., the traders comprising the market) are represented by mutually interacting spins and the arrival of information by external fields, have also been used to simulate the financial market [19–22]. Among nonmicroscopic approaches, multifractal processes have been used extensively for modeling such scale invariant properties [23, 24]. The multifractal random walk model has generalized the usual random walk model of financial price changes and accounts for many of the observed empirical properties [25].

However, on the empirical front, there is some controversy about the universality of the power-law nature for the tails of the index return distribution. In the case of developed markets, for example the All Ordinaries index of the Australian stock market, the negative tail has been reported to follow the inverse cubic law while the positive tail is closer to Gaussian [26]. Again, other studies of the Hang Seng and Nikkei indices report the return distribution to be exponential [27, 28]. For developing economies, the situation is even less clear. There have been several claims that emergent markets have a return distribution that is significantly different from developed markets. For example, a recent study contrasting the behavior of indices from seven developed markets with the KOSPI index of the Korean stock market found that while the former exhibit the inverse cubic law, the latter follows an exponential distribution [17]. Another study of the Korean stock market reported that the index distribution has changed to exponential from a power-law nature only in recent years [29]. On the other hand, the IBOVESPA index of the Sao Paulo stock market has been claimed to follow a truncated Levy distribution [30, 31]. However, there have also been reports of the inverse cubic law for emerging markets, for example for the Mexican stock market index IPC [32] and the WIG20 index of the

Polish stock market [33]. A comparative analysis of 27 indices from both mature and emerging markets found their tail behavior to be similar [34]. It is of course difficult to conclude about the nature of the fluctuation distribution for individual stock prices from the index data, as the latter is a weighted average of several stocks. Therefore, in principle, the index can show a distribution quite different from that of its constituent stocks if their price movements are not correlated.

Many of the studies reported above have only used graphical fitting to determine the nature of the observed return distribution. This has recently come under criticism as such methods often result in erroneous conclusions. Hence, a more accurate study using reliable statistical techniques needs to be carried out to decide whether emerging markets do behave similarly to developed markets in terms of fluctuations. Here, we will show how to carry out such an analysis using data from the Indian financial markets. The Indian data is very important for deciding whether emerging markets behave differently from developed markets, as it is one of the fastest growing financial markets in the world. A recent study of individual *stock prices* in the National Stock Exchange (NSE) of India has claimed that the corresponding return distribution is exponentially decaying at the tails [35], and not the inverse cubic law that is observed for developed markets [13, 14]. However, a more detailed study over a larger dataset has established the inverse cubic law for individual *stock prices* [36]. On the other hand, to get a sense of the nature of fluctuations for the entire market, one needs to look at the corresponding distribution for the market index. Although the individual stock prices and the market index are related, it is not obvious that they should have the same kind of distribution, as this relation is dependent on the degree of correlation between different stock price movements.

We will show in the following how to use both individual price as well as overall market index data to establish the nature of the distribution. Although we use the specific example of Indian markets, it is needless to say that the analysis applies to other markets as well.

5.1.2
Stock Price Return Distribution

Most early studies of stock price fluctuations in emerging markets were done on low-frequency daily data. Let us instead focus on high-frequency tick-by-tick data, which will be complemented by analysis of daily data over much longer periods. The dataset that we have chosen for this purpose is from the National Stock Exchange (NSE) of India, the largest among the 23 exchanges in India, with more than 85% of the total value of transactions for securities in all market segments of the entire Indian financial market in recent times [37]. This data set is of unique importance, as we have access to daily data right from the time the market commenced operations in the equities market in November 1994, up to the present when it has become the world's third largest stock exchange (after NASDAQ and NYSE) in terms of transactions [38]. Over this period, the market has grown rapidly, with the number of transactions having increased by more than three orders of magnitude. Therefore, if markets do show discernible transition in the return distribution during their evolution, the Indian market data is best placed to spot evidence for it, not least because of the rapid transformation of the Indian economy in the liberalized environment since the 1990s.

We focus on two important questions: (i) Does the market exhibit a price fluctuation different from the inverse cubic law form seen in developed markets, and (ii) if the market is indeed following the inverse cubic law at present, whether this has been converged, starting from an initially different distribution when the market had just begun operation. Both of these questions are answered negatively in the following analysis.

The two datasets having different temporal resolutions that we analyze are the following. (i) The high-frequency tick-by-tick data contains information about all transactions carried out in the NSE between January 2003 and March 2004. This information includes the date and time of trade, the price of the stock during transaction and the volume of shares traded. This database is available in the form of CDs published by the NSE. For calculating the price return, we have focused on 489 stocks that were part of the BSE 500 index (a comprehensive indicator for the Indian financial market) during this period. The number of transactions for each company in this set is $\sim 10^6$, on the average. The total number of transactions for the 489 stocks is on the order of 5×10^8 during the period under study. (ii) The daily closing price of all the stocks listed in NSE during its period of existence between November 1994 and May 2006. This was obtained from the NSE website [39] and manually corrected for stock splitting.[1] For comparison with US markets, in particular the NYSE, we consider the 500 stocks listed in S & P 500 during the period November 1994 to May 2006, the daily data being obtained from Yahoo! Finance [40].

To measure the price fluctuations such that the result is independent of the scale of measurement, we calculate the logarithmic return of price. If $P_i(t)$ is the stock price of the ith stock at time t, then the (logarithmic) price return is defined as

$$R_i(t, \Delta t) \equiv \ln P_i(t + \Delta t) - \ln P_i(t) \,. \tag{5.3}$$

However, the distribution of price returns of different stocks may have different widths, owing to differences in their volatility, defined (for the ith stock) as $\sigma_i^2 \equiv \langle R_i^2 \rangle - \langle R_i \rangle^2$. To compare the distribution of different stocks, we normalize the returns by dividing them with their volatility $\sigma_i(t)$, as in [4]. The resulting normalized

1) When the price of a stock increases to such levels that trading in it reduces significantly, the company often decides to split a single share into multiple parts in order to increase liquidity. In the stock price data we are analyzing here, this may be reflected as a sudden drop in the price. As it is not related to any change in the intrinsic value of the stock, this should be corrected in the data so as to avoid obtaining an erroneous return distribution.

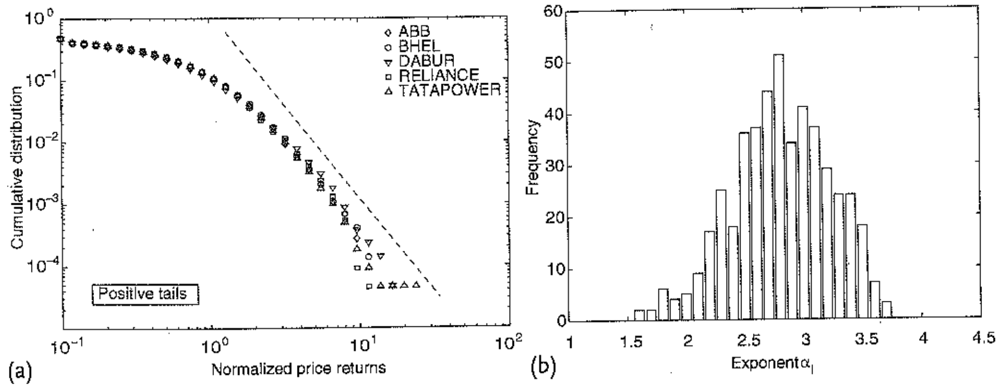


Figure 5.1 (a) Cumulative distribution of the positive tails of the normalized 5 min returns distribution of 5 stocks chosen arbitrarily from those listed in the NSE for the period January 2003 to March 2004. The broken line indicates a power law with exponent $\alpha = 3$. (b) The histogram of the power-law exponents obtained by regression fit for the positive tail of individual cumulative return distributions of 489 stocks. The median of the exponent values is 2.84.

price return[2)] is given by

$$r_i(t, \Delta t) \equiv \frac{R_i(t) - \langle R_i(t) \rangle}{\sigma_i(t)}, \tag{5.4}$$

where $\langle \cdots \rangle$ denotes the time average over the given period.

For analysis of the high-frequency data, we consider the aforementioned 489 stocks. Choosing an appropriate Δt, we obtain the corresponding return by taking the log ratio of consecutive average prices, averaged over a time window of length Δt. Figure 5.1a shows the cumulative distribution of the normalized returns r_i with $\Delta t = 5$ min for five stocks, arbitrarily chosen from the dataset. We observe that the distribution of normalized returns r_i for all the stocks have the same functional form with a long tail that follows a power-law asymptotic behavior. The distribution of the corresponding power-law exponent α_i for all the 489 stocks that we have considered is shown in Figure 5.1b.

As all the individual stocks follow very similar distributions, we can merge the data for different stocks to obtain a single distribution for normalized returns. The aggregated return dataset with $\Delta t = 5$ min has 6.5×10^6 data points. The corresponding cumulative distribution is shown in Figure 5.2a, with the exponents for the positive and negative tails estimated as

$$\alpha = \begin{cases} 2.87 \pm 0.08 & \text{(positive tail)} \\ 2.52 \pm 0.04 & \text{(negative tail)} \end{cases}. \tag{5.5}$$

From this figure we confirm that the distribution does indeed follow a power-law decay, albeit with different exponents for the positive and negative return tails.

2) The normalization of return $R_i(t)$ is performed by removing its own contribution from the volatility, that is $\sigma_i(t) = \sqrt{\frac{1}{N-1}\sum_{t' \neq t}\left[R_i(t') - \langle R_i(t)\rangle\right]^2}$.

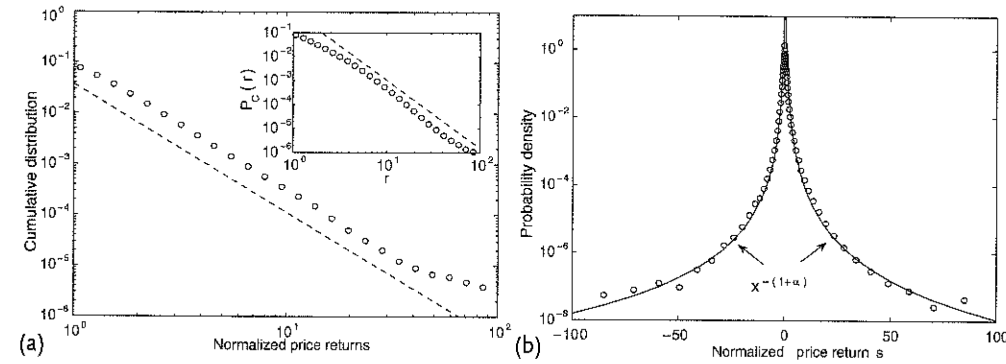


Figure 5.2 (a) Cumulative distribution of the negative and (inset) positive tails of the normalized returns for the aggregated data of 489 stocks in the NSE for the period January 2003 to March 2004. The broken line for visual guidance indicates the power-law asymptotic form. (b) Probability density function of the normalized returns. The solid curve is a power-law fit in the region 1–50. We find that the corresponding cumulative distribution exponent, $\alpha = 2.87$ for the positive tail and $\alpha = 2.52$ for the negative tail.

Such a difference between the positive and negative tails have also been observed in the case of stocks in the NYSE [14]. To further verify that the tails are indeed consistent with a power-law form, we perform an alternative measurement of α using the Hill estimator [43, 44]. We arrange the returns in decreasing order such that $r_1 > \cdots > r_n$ and obtain the Hill estimator (based on the largest $k+1$ values) as $H_{k,n} = \frac{1}{k}\sum_{i=1}^{k}\log\frac{r_i}{r_{k+1}}$, for $k = 1, \cdots, n-1$. The estimator $H_{k,n} \to \alpha^{-1}$ when $k \to \infty$ and $k/n \to 0$. For our data, this procedure gives $\alpha = 2.86$ and 2.56 for the positive and the negative tail, respectively (when $k = 20\,000$), which are consistent with (5.5).

Next, we extend this analysis for longer time-scales, to observe how the nature of the distribution changes with increasing Δt. As has been previously reported for US markets, the distribution is found to decay faster as Δt becomes large. However, up to $\Delta t = 1$ day, that is the daily closing returns, the distribution clearly shows a power-law tail (Figure 5.3a). The deviation is because of the decreasing size of the dataset with increase in Δt. Note that, while for $\Delta t < 1$ day we have used the high-frequency data, for $\Delta t = 1$ day we have considered the longer dataset of closing price returns for all stocks traded in NSE between November 1994 to May 2006. In Figure 5.3b we have also shown the distributions of the power-law exponents for the individual stocks, for time-scales varying between 10 min $\leq \Delta t \leq$ 60 min. We observe that the bulk of the exponents fall between 2 and 4, consistent with the results from the merged datasets.

To compare the distribution of returns in this emerging market with that observed in mature markets, we have considered the daily return data for the 500 stocks from NYSE listed in S & P 500 over the same period. As seen in Figure 5.4, the distributions for NSE and NYSE are almost identical, implying that the price fluctuation distribution of emerging markets cannot be distinguished from that of developed markets, contrary to what has been claimed recently [35].

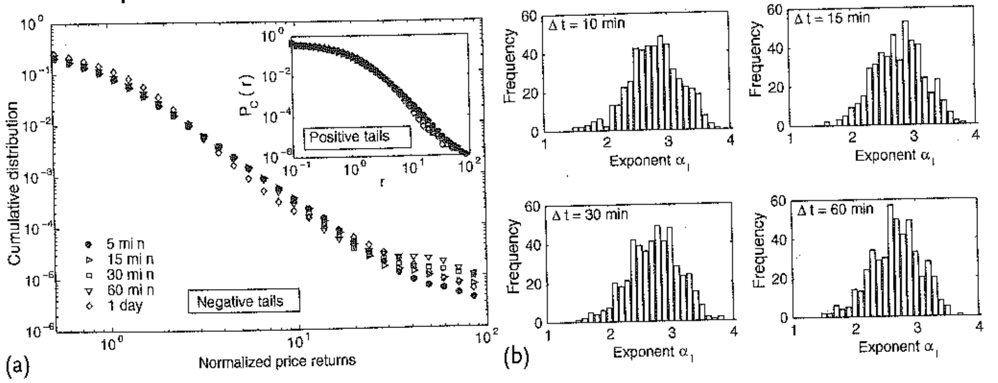


Figure 5.3 (a) Cumulative distribution of the negative and (inset) positive tails of the normalized returns distribution for different time-scales ($\Delta t \leq 1$ day). (b) Histograms of the power-law exponents for each of the 489 stocks, obtained by regression fit on the positive tail of cumulative return distributions, for different time-scales (10 min $\leq \Delta t \leq$ 60 min).

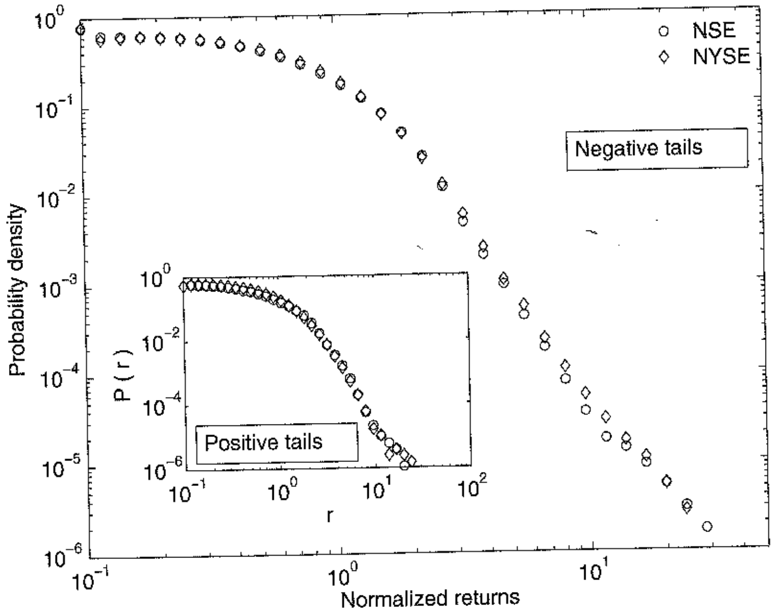


Figure 5.4 Comparison of the negative and (inset) positive tails of the normalized daily returns distribution for all stocks traded at NSE (○) and 500 stocks traded at NYSE (◇) during the period November 1994 to May 2006.

We now turn to the second question, and check whether it is possible to see any discernible change in the price fluctuation distribution as the stock market evolved over time. For this we focus on the daily return distribution for *all* stocks that were traded during the entire period of existence of the NSE. This period is divided into four intervals (a) 1994–1996, (b) 1997–1999, (c) 2000–2002, and (d) 2003–2006 [36],

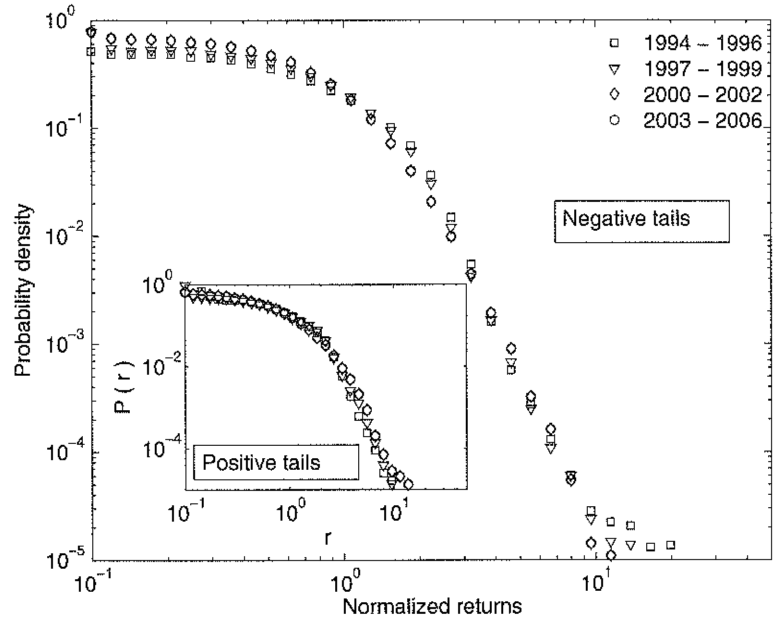


Figure 5.5 The negative and (inset) positive tails of the normalized daily returns distribution for all NSE stocks traded during the periods 1994–1996 (□), 1997–1999 (▽), 2000–2002 (◇) and 2003–2005 (○).

each corresponding to an increase in the number of transactions by an order of magnitude. Figure 5.5 shows that the return distribution at all four periods are similar, the negative tail even more so than the positive one. While the numerical value of the tail exponent may appear to have changed somewhat over the period that the NSE has operated, the power-law nature of the tail is apparent at even the earliest period of its existence. We therefore conclude that the convergence of the return distribution to a power-law functional form is extremely rapid, indicating that a market is effectively always at the nonequilibrium steady state characterized by the inverse cubic law.

We have also verified that stocks in the Bombay Stock Exchange (BSE), the second largest in India after NSE, follow a similar distribution [41]. Moreover, the return distribution of several Indian market indices (e.g., the NSE Nifty) also exhibit power-law decay, with exponents very close to 3 [42]. As the index is a composite of several stocks, this behavior can be understood as a consequence of the power-law decay for the tails of individual stock price returns, provided the movement of these stocks are correlated [14, 41]. Even though the Indian market microstructure has been refined and modernized significantly in the period under study as a result of the reforms and initiatives taken by the government, the nature of the return distribution has remained invariant, indicating that the nature of price fluctuations in financial markets is most probably independent of the level of economic development.

Why did previous studies miss a long tail in the distribution of stock price returns in the Indian market? Paucity of data can result in missing the long tail of a power-law distribution and falsely identifying it to be an exponential distribution. Matia *et al.* [35] claimed that differences in the daily return distribution for Indian and US ma rkets were apparent even if one looks at only 49 stocks from each market. However, this statement is critically dependent upon the choice of stocks. Indeed, when we make an arbitrary choice of 50 stocks in both Indian and US markets, and compare their distributions, we find them to be indistinguishable. Therefore, the results of analysis done on such small datasets can hardly be considered stable, with the conclusions depending on the particular sample of stocks.

5.1.3 Market Index Return Distribution

We will now focus on the two largest stock exchanges in India, the NSE and the Bombay Stock Exchange (BSE). NSE, the more recent of the two, is not only the most active stock exchange in India, but also the third largest in the world in terms of transactions [45]. As already mentioned earlier, we shall study the behavior of this market over the entire period of its existence. During this period, the NSE has grown by several orders of magnitude (Figure 5.6) demonstrating its emerging character. In contrast, BSE is the oldest stock exchange in Asia, and was the largest in India until the creation of NSE. However, over the past decade its share of the Indian financial market has fallen significantly. Therefore, we contrast two markets which have evolved very differently in the period under study.

In this subsection, we show that the Indian financial market, one of the largest emerging markets in the world, has index fluctuations similar to that seen for developed markets. Further, we find that the nature of the distribution is invariant with respect to different market indices, as well as the time-scale of observation. Taken together with the results of the previous subsection on the distribution of individual stock price returns in Indian markets, this strongly argues in favor of the universality of the nature of fluctuation distribution, regardless of the stage of development of the market or the economy underlying it.

The primary dataset that is used here is that of the Nifty index of the NSE which, along with the Sensex of BSE, is one of the primary indicators of the Indian market. It is composed of the top 50 highly liquid stocks which make up more than half of the market capitalization in India. We have used (i) high frequency data from January 2003 to March 2004, where the market index is recorded every time a trade takes place for an index component. The total number of records in this database is about 6.8×10^7. We have also looked at data over much longer periods by considering daily closing values of (ii) the Nifty index for the 16-year period July 1990 to May 2006, and (iii) the Sensex index of BSE for the 15-year period January 1991 to May 2006. In addition, we have also looked at the BSE 500 index for the much shorter period February 1999 to May 2006. Sensex consists of the 30 largest and most actively traded stocks, representative of various sectors of BSE, while the BSE 500 is calculated using 500 stocks representing all 20 major sectors of the economy.

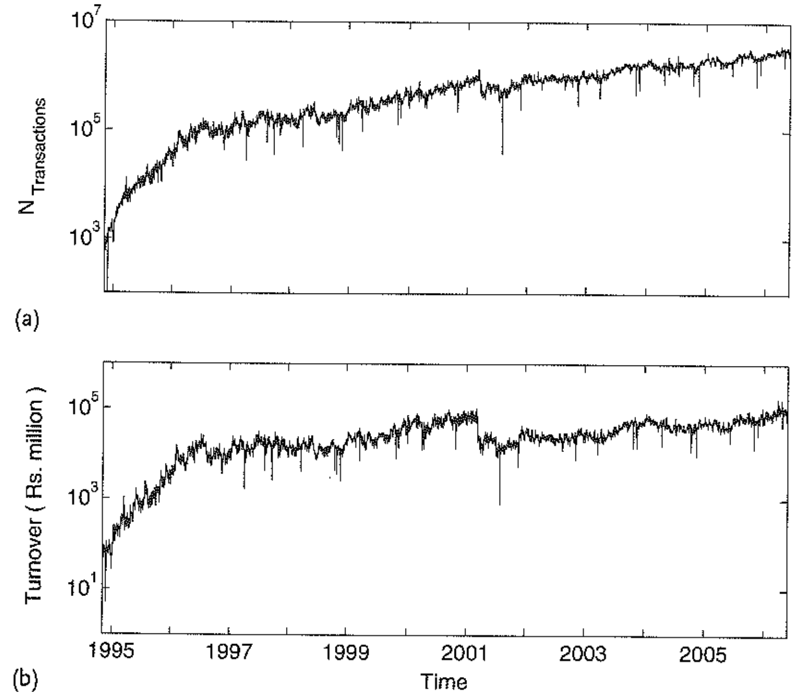


Figure 5.6 Time-evolution of the National Stock Exchange of India from 1994–2006 in terms of (a) the total number of trades and (b) the total turnover (i.e., traded value).

We first analyze the high-frequency data for the NSE Nifty index, which we sampled at 1-min intervals to generate the time-series $I(t)$. From $I(t)$ we compute the logarithmic return $R_{\Delta t}(t)$, defined in (5.1). These return distributions calculated using different time intervals may have varying width, owing to differences in their volatility, defined as $\sigma^2_{\Delta t} \equiv \langle R^2 \rangle - \langle R \rangle^2$, where $\langle \ldots \rangle$ denotes the time average over the given time period. Hence, to be able to compare the distributions, we need to normalize the returns $R(t)$ by dividing them by the volatility $\sigma_{\Delta t}$. However, this leads to systematic underestimation of the tail of the normalized return distribution. This is because, even when a single return $R(t)$ is very large, the scaled return is bounded by $\sqrt{N}$, as the same large return also contributes to the variance $\sigma_{\Delta t}$. To avoid this, we remove the contribution of $R(t)$ itself from the volatility, and the new rescaled volatility is defined as

$$\sigma_{\Delta t}(t) = \sqrt{\frac{1}{N-1} \sum_{t' \neq t} \{R(t', \Delta t)\}^2 - \langle R(t', \Delta t) \rangle^2} \,, \tag{5.6}$$

as described in [4]. The resulting *normalized* return is given by,

$$r(t, \Delta t) \equiv \frac{R - \langle R \rangle}{\sigma_{\Delta t}(t)} \,. \tag{5.7}$$

Prior to obtaining numerical estimates of the distribution parameters, we carry out a test for the nature of the return distribution, that is whether it follows a power law or an exponential or neither. For this purpose we use a statistical tool that is independent of the quantitative value of the distribution parameters. Usually, it is observed that the tail of the return distribution decays at a slower rate than the bulk. Therefore, the determination of the nature of the tail depends on the choice of the lower cut-off u of the data used for fitting a theoretical distribution. To observe this dependence on the cut-off u, we calculate the TP and TE statistics [46, 47] as a function of u, comparing the behavior of the tail of the empirical distributions with power law and exponential functional forms, respectively. These statistics converge to zero if the underlying distribution follows a power law (TP) or exponential (TE), regardless of the value of the exponent or the scale parameter. On the other hand, they deviate from zero if the observed return distribution differs from the target theoretical distribution (power law for TP and exponential for TE).

5.1.3.1 TP Statistic

Consider the power-law distribution

$$F(x) = 1 - P_c(x) = 1 - (u/x)^{\alpha}\,, \quad \text{for} \quad x \geq u\,, \tag{5.8}$$

where u is the lower cut-off, and α is the power-law exponent for the distribution. For a finite sample $x_1, \ldots, x_n$, the TP statistic, $\mathrm{TP}(x_1, \ldots, x_n)$, is defined such that it converges to zero asymptotically for large n [46, 47]. If the underlying distribution for a sample differs from the power-law form given in (5.8), TP is seen to deviate from zero. This statistic is based on the first two normalized statistical log-moments of the power-law distribution

$$E_1 = E\left[\log\frac{X}{u}\right] = \int_u^{\infty} \log\frac{x}{u}\, d\,F(x) = \frac{1}{\alpha}\,, \tag{5.9}$$

and

$$E_2 = E\left[\log^2\frac{X}{u}\right] = \int_u^{\infty} \log^2\frac{x}{u}\, d\,F(x) = \frac{2}{\alpha^2}\,, \tag{5.10}$$

where, $E[z]$ represents the mathematical expectation of z. The TP statistic is then defined as

$$\mathrm{TP} = \left[\frac{1}{n}\sum_{k=1}^{n}\log\frac{x_k}{u}\right]^2 - \frac{1}{2n}\sum_{k=1}^{n}\log^2\frac{x_k}{u}\,, \tag{5.11}$$

which tends to zero as $n \to \infty$. The estimation of the standard deviation for the TP statistic is provided by the standard deviation of the sum

$$\left(\frac{E_2}{2} - 2E_1^2\right) + \frac{1}{n}\sum_{k=1}^{n}\left[2E_1\log\frac{x_k}{u} - \frac{1}{2}\log^2\frac{x_k}{u}\right]. \tag{5.12}$$

5.1.3.2 TE Statistic

Consider the exponential distribution

$$F(x) = 1 - P_c(x) = 1 - \exp(-(x-u)/d)\,, \quad \text{for} \quad x \geq u\,, \tag{5.13}$$

where u is the lower cut-off, and $d(> 0)$ is the scale parameter of the distribution. For a finite sample $x_1, \ldots, x_n$, the TE statistic, $\mathrm{TE}(x_1, \ldots, x_n)$, is defined such that it converges to zero asymptotically for large n [46]. If the underlying distribution for a sample differs from the exponential form given in (5.13), TE is seen to deviate from zero. This statistic is based on the first two normalized statistical (shifted) log-moments of the exponential distribution

$$\begin{aligned} E_1 &= E\left[\log\left(\frac{X}{u} - 1\right)\right] = \int_u^{\infty} \log\left(\frac{x}{u} - 1\right) d\,F(x) \\ &= \log\frac{d}{u} - \gamma\,, \end{aligned} \tag{5.14}$$

where $\gamma = 0.577215$ is the Euler constant, and

$$\begin{aligned} E_2 &= E\left[\log^2\left(\frac{X}{u} - 1\right)\right] = \int_u^{\infty} \log^2\left(\frac{x}{u} - 1\right) d\,F(x) \\ &= \left(\log\frac{d}{u} - \gamma\right)^2 + \frac{\pi^2}{6}\,. \end{aligned} \tag{5.15}$$

As before, $E[\ldots]$ denotes the mathematical expectation. The TE statistic is then defined as

$$\mathrm{TE} = \frac{1}{n}\sum_{k=1}^{n}\log^2\left(\frac{x_k}{u} - 1\right) - \left[\frac{1}{n}\sum_{k=1}^{n}\log\left(\frac{x_k}{u} - 1\right)\right]^2 - \frac{\pi^2}{6}\,, \tag{5.16}$$

which tends to zero as $n \to \infty$. The estimation of the standard deviation for the TE statistic is provided by the standard deviation of the sum

$$\frac{1}{n}\sum_{k=1}^{n}\left[\log\left(\frac{x_k}{u} - 1\right) - E_1\right]^2. \tag{5.17}$$

Figure 5.7 shows visually the deviation of the empirical data from the power law and exponential distributions. The TP and the TE statistics are plotted as functions of the lower cut-off u for 1-min returns of the NSE Nifty index. The TP statistic shows a large deviation till $u \leq 1$, after which it converges to zero indicating power-law behavior for large u. Correspondingly, the TE statistic excludes an exponential model for $u \geq 1$ as well as for very low values of u, although over the intermediate range $2 \times 10^{-1} < u < 6 \times 10^{-1}$ an exponential approximation may be possible.

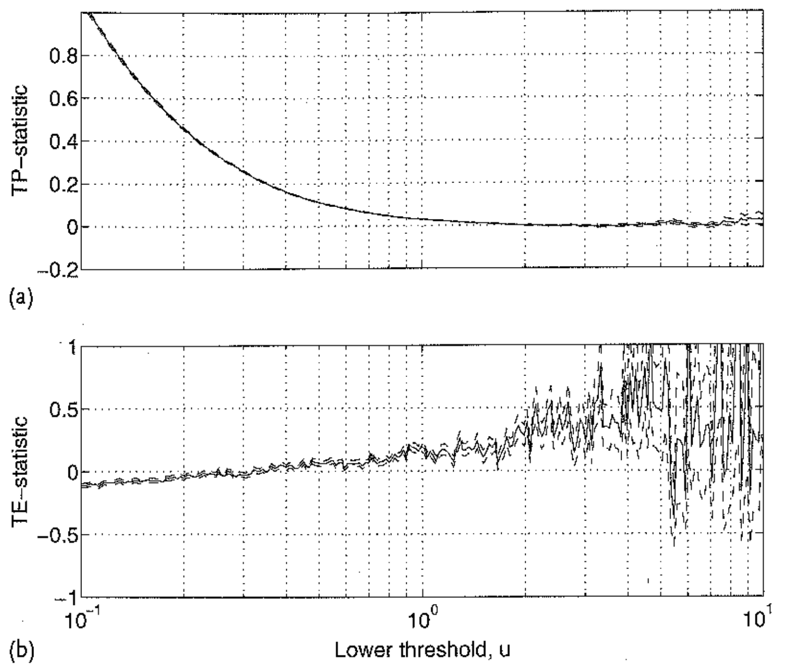


Figure 5.7 (a) TP statistic and (b) TE statistic as a function of the lower cut-off u for positive returns of the NSE with time interval $\Delta t = 1$ min. The broken lines indicate plus or minus one standard deviation of the statistics.

Figure 5.8 shows the cumulative distribution of the normalized returns for $\Delta t =$ 1 min. For both positive and negative tails, there is an asymptotic power-law behavior. The power-law regression fit for the region $r \geq 2$ give exponents for the positive and the negative tails estimated as

$$\alpha = \begin{cases} 2.98 \pm 0.09 & \text{(positive tail)} \\ 3.37 \pm 0.10 & \text{(negative tail)} \end{cases} . \tag{5.18}$$

Note that, to avoid artifacts due to the data measurement process in the calculation of return distribution for $\Delta t < 1$ day, we have removed the returns corresponding to overnight changes in the index value.

We also perform an alternative estimation of the tail index of the above distribution by using the Hill estimator [43], which is the maximum likelihood estimator of α. For finite samples, however, the expected value of the Hill estimator is biased and depends crucially on the choice of the number of order statistics used for calculation. We have used the bootstrap procedure [48] to reduce this bias and to choose the optimal number of order statistics for calculating the Hill estimator, described in detail below. We found $\alpha \simeq 3.22$ and 3.47 for the positive and the negative tails, respectively.

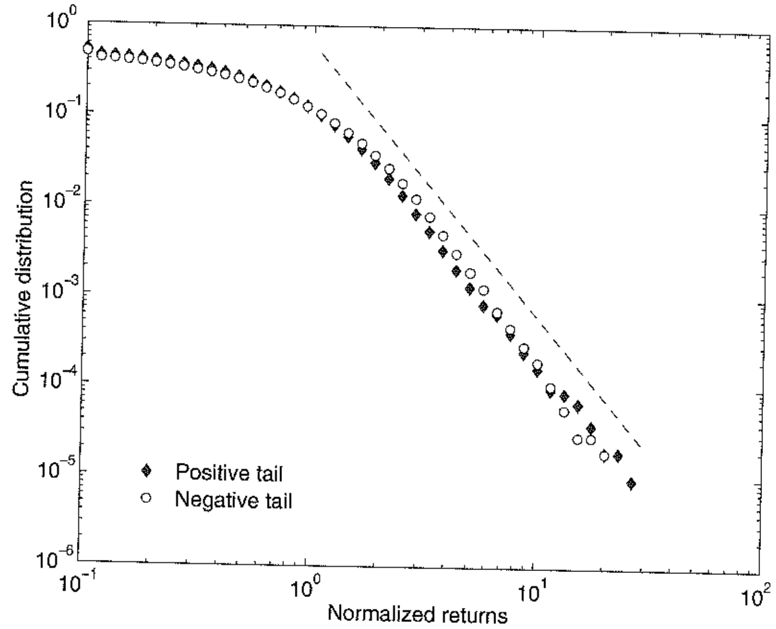


Figure 5.8 The cumulative distribution of the normalized 1-min return for the NSE Nifty index. The broken line indicates a power law with exponent $\alpha = 3$.

5.1.3.3 Hill Estimation of Tail Exponent

The Hill estimator gives a consistent estimate of the tail exponent α from random samples of a distribution with an asymptotic power-law form. First, we arrange the returns in decreasing order such that $r_1 > \cdots > r_n$. Then the Hill estimator (based on the largest $k + 1$ values) is given as

$$\gamma_{k,n} = \frac{1}{k} \sum_{i=1}^{k} \log \frac{r_i}{r_{k+1}} , \tag{5.19}$$

for $k = 1, \cdots, n-1$. The estimator $\gamma_{k,n} \to \alpha^{-1}$ when $k \to \infty$ and $k/n \to 0$. However, for a finite time-series, the expectation value of the Hill estimator is biased, that is it will consistently over or underestimate α. Further, γ depends critically on our choice of k, the order statistics used to compute the Hill estimator.

If the form of the distribution function from which the random sample is chosen is known, then the bias and the stochastic error variance of the Hill estimator can be calculated. From this, the optimum k value can be obtained such that the asymptotic mean-square error of the Hill estimator is minimized. Increasing k reduces the variance because more data are used, but increases the bias because the power law is assumed to hold only in the extreme tail. Unfortunately, the distribution for the empirical data is not known and hence this procedure has to be replaced by an asymptotically equivalent data driven process.

One such method is the subsample bootstrap method. This method can be used to estimate an optimal number for the order statistics ($\bar{k}$) that will reduce the asymptotic mean-square error of the Hill estimator. However, this process requires the choice of certain parameters, for example the subsample size n_s and the range of k values in which one searches for the minimum of the bootstrap statistics. We briefly describe this procedure below; for details and mathematical validation of this procedure, please see [48].

We assume the underlying empirical distribution function to be heavy-tailed, viz.,

$$P_c(x) = ax^{-\alpha}\left[1 + bx^{-\beta} + o(x^{-\beta})\right], \tag{5.20}$$

with $\alpha, \beta, a > 0$ and $-\infty < b < \infty$. We first calculate an initial $\gamma_0 = \gamma_{k_0,n}$ for the original series with a reasonably chosen (but nonoptimal) k_0. Then we choose various subsamples of size n_s randomly from the original series, which are orders of magnitude smaller than n. The quantity γ_0 is a good approximation of subsample α^{-1}, since the error in γ is much larger for n_s than for n observations. The optimal order statistics $\bar{k}_s$ for the subsample is found by computing $\gamma(k_s, n_s)$ for different values of k_s and then minimizing the deviation from γ_0. Given $\bar{k}_s$, the suitable full sample $\bar{k}$ can be found by using

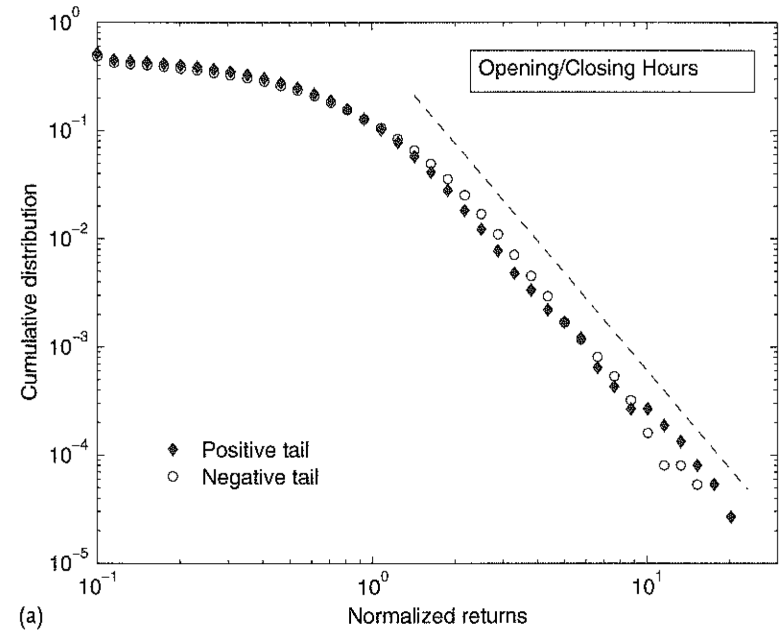
$$\bar{k} = \bar{k}_s\left(\frac{n}{n_s}\right)^{\frac{2\beta}{2\beta+\alpha}} . \tag{5.21}$$

Here the initial estimate of α is taken to be $1/\gamma_0$. Further, we have considered $\beta = \alpha$, as done by Hall [49], although the results are not very sensitive to the choice of β. Once $\bar{k}$ is calculated, the final estimate of the tail index is given by $\alpha = 1/\gamma_{\bar{k},n}$.

For calculating the initial γ_0 we have chosen k_0 to be 0.5% of the sample size n. One thousand subsamples, each of size $n_s = n/40$, are randomly picked from the full dataset. To obtain optimal k_s, we confine ourselves to 4% of the subsample size n_s. To find the stochastic error in our estimation of α, we have computed the 95% confidence interval as given by $\pm 1.96[1/(\alpha^2 m)]^{1/2}$. Although a Jackknife algorithm can also be used to calculate this error bound, the results obtained using this method will be close to that obtained using the bootstrap method over many realizations [48], as we have done here.

5.1.3.4 Temporal Variations in the Return Distribution

To investigate the effect of *intra-day* variations in market activity, we analyze the 1-min return time-series by dividing it into two parts, one corresponding to returns generated in the opening and the closing hours of the market, and the other corresponding to the intermediate time period. In general, it is known that the average intra-day volatility of stock returns follows a U-shaped pattern [50, 51] and one can expect this to be reflected in the nature of the fluctuation distribution for the opening and closing periods, as opposed to the intervening period. We indeed find the index fluctuations for these two datasets to be different (Figure 5.9). In particular, the cumulative distribution tail for the opening and closing hour returns show a power-law scaling with exponent close to 3, whereas for the intermediate period we



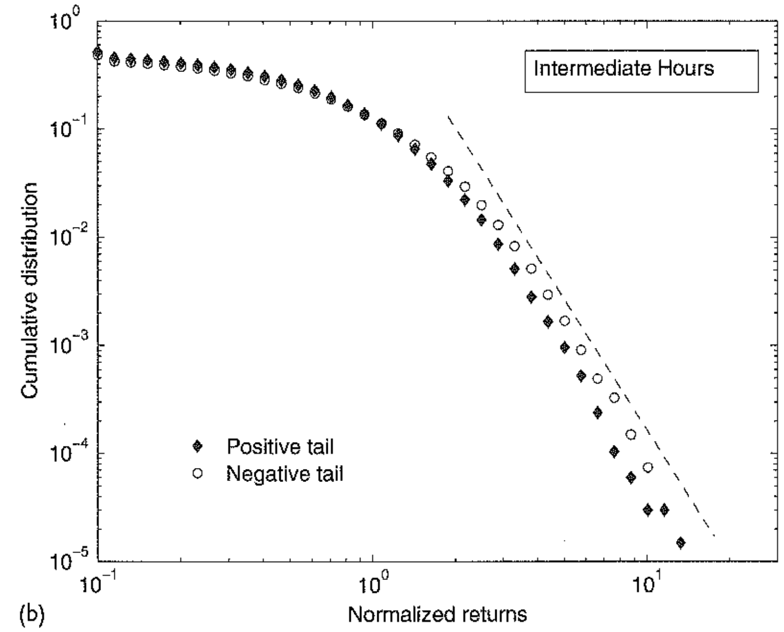


Figure 5.9 Intra-day variation in the cumulative distribution of the normalized 1-min return for the NSE Nifty index: return distribution during (a) the opening and closing hours (the broken line indicates a power law with exponent $\alpha = 3$) and (b) the intermediate time period (the broken line indicates a power law with exponent $\alpha = 4$).

see that the exponent is close to 4. This observation is similar to that reported for the German DAX index, where removal of the first few minutes of return data after the daily opening resulted in a power-law distribution with a different exponent compared to the intact dataset [52].

Next, we extend our analysis for longer time-scales, Δt. We find that time aggregation of the data increases the α value. The tail of the return distribution still retains its power-law form (Figure 5.10), until at longer time-scales the distribution slowly converges to Gaussian behavior (Table 5.1). The results are invariant with respect to whether one calculates return using the sampled index value at the end point of an interval or the average index value over the interval. Figure 5.10 shows the cumulative distribution of normalized Nifty returns for time-scales up to 60 min. However, using a similar procedure for generating *daily* returns from the tick-by-tick data would give us a very short time-series. This is not enough for reliable analysis as it takes at least 3000 data points for a meaningful estimate of the tail index. For this reason, we have analyzed the daily data using a different source, with the time period stretching over a considerably longer period (16 years). The return distribution of the daily closing data of Nifty shows qualitatively similar behavior to the 1 min distribution. The Sensex index, which is from another stock exchange, also follows a similar distribution (Figure 5.11). The measured exponent values are all close to 3. This does not contradict the earlier observation that α increases with Δt, because, increasing the sample size (as has been done

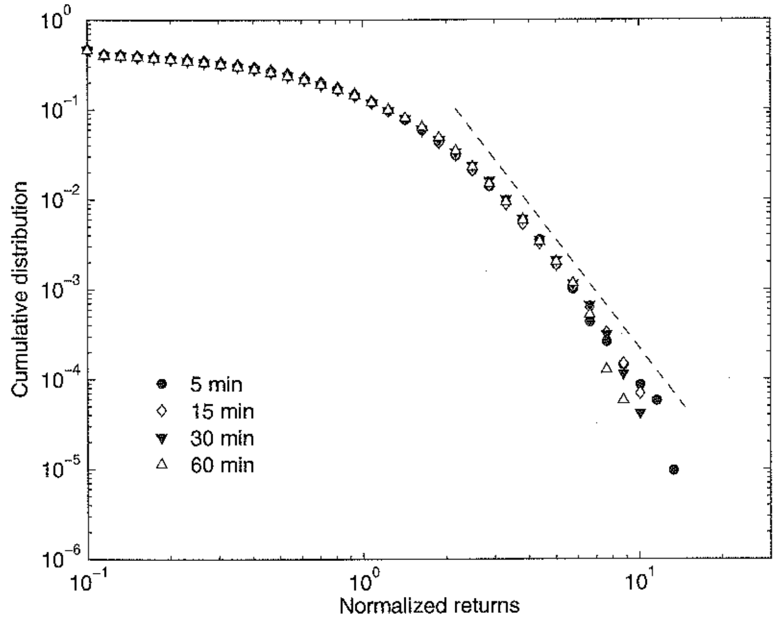


Figure 5.10 The negative tail of the cumulative distribution of the NSE Nifty index returns for different time intervals Δt up to 60 min. The broken line indicates a power law with exponent $\alpha = 4$.

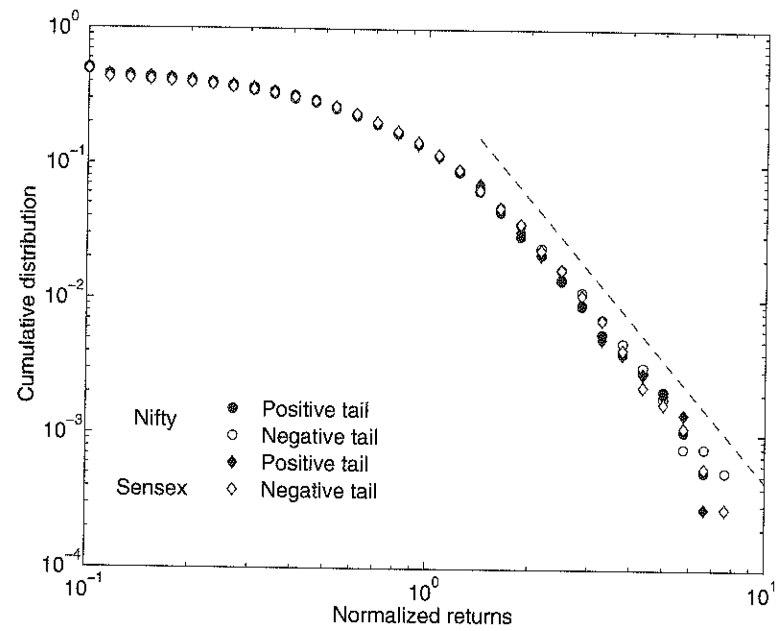


Figure 5.11 The cumulative distribution of the normalized 1-day return for the NSE Nifty and BSE Sensex index. The broken line indicates a power law with exponent $\alpha = 3$.

Table 5.1 Comparison of the power-law exponent α of the cumulative distribution function for various index returns. Power-law regression fits are done in the region $r \geq 2$. The Hill estimator is calculated using the bootstrap algorithm.

Index	Δt	Power-law fit		Hill estimator	
		Positive	Negative	Positive	Negative
Nifty (2003–2004)	1 min	2.98 ± 0.09	3.37 ± 0.10	3.22 ± 0.03	3.47 ± 0.03
	5 min	4.42 ± 0.37	3.44 ± 0.21	4.51 ± 0.03	4.84 ± 0.03
	15 min	5.58 ± 0.88	3.96 ± 0.27	6.25 ± 0.03	4.13 ± 0.04
	30 min	5.13 ± 0.41	3.92 ± 0.45	5.65 ± 0.03	4.30 ± 0.03
	60 min	5.99 ± 1.52	4.42 ± 0.65	7.85 ± 0.03	5.11 ± 0.04
Nifty (1990–2006)	1 day	3.10 ± 0.34	3.18 ± 0.28	3.33 ± 0.14	3.37 ± 0.14
Sensex (1991–2006)	1 day	3.33 ± 0.77	3.45 ± 0.25	2.93 ± 0.15	3.84 ± 0.12

for $\Delta t = 1$ day) improves the estimation of α. This underlines the invariance of the nature of market fluctuations with respect to time aggregation, interval used and different exchanges.

The much shorter dataset of the BSE 500 daily returns shows a significant departure from power-law behavior, essentially following an exponential distribution. This is not surprising, as looking at data over shorter periods can result in misidentification of the nature of the distribution. Specifically, the relatively low number of data points corresponding to returns of large magnitude can lead to missing out the long tail. In fact, even for individual stocks in developed markets, although the tails follow a power law, the bulk of the return distribution is exponential [53]. This problem arising from using limited datasets might be one of the reasons why some studies have seen significant deviation of index return distribution from a power law.

A more serious problem is that the analysis in many of these studies is usually performed only by graphically fitting the data with a theoretical distribution function. Such a visual judgement of the goodness of fit may lead to erroneous characterization of the nature of fluctuation distribution. Graphical procedures are often subjective, particularly with respect to the choice of the lower cut-off up to which fitting is carried out. This dependence of the theoretical distribution that best describes the tail on the cut-off, has been explicitly demonstrated through the use of TP and TE statistics above. Moreover, recent studies have criticized the reliability of graphical methods by showing that least square fitting for estimating the power-law exponent tends to provide biased estimates, while the maximum likelihood method produces more accurate and robust estimates [54, 55]. It is for this reason that we have used the Hill estimator to determine the tail exponents.

If the individual stocks follow the inverse cubic law, it would be reasonable to suppose that the index, which is a weighted average of several stocks, will also behave similarly, provided the different stocks move in a correlated fashion [16]. As the price movements of stocks in an emerging market are even more correlated than in developed markets [56], it is expected that the returns for stock prices and the index should follow the same distribution. Therefore, the demonstration of the inverse cubic law for the index fluctuations in the Indian market is consistent with our preceding demonstration [36] showing that the individual stock prices in this market follow the same behavior.

On the whole, our study points out the remarkable robustness of the nature of the fluctuation distribution for markets. While, in the period under study, the NSE had begun operation and rapidly increased in terms of activity, the BSE had existed for a long time prior to this period and showed a significant decrease in market share. However, both showed very similar fluctuation behavior. This indicates that, at least in the Indian context, the distribution of returns is invariant with respect to markets. The fact that the distribution is quantitatively the same as in developed markets, implies that it is also probably independent of the state of the economy. In addition, our observation that the intra-day return distribution of Indian market index shows properties similar to that reported for developed markets, suggests that even at this level of detail the fluctuation behavior of the two kinds of markets are rather similar. Therefore, our results indicate that although markets may differ from each other in terms of (i) the details of their components, (ii) the nature of interactions and (iii) their susceptibility to news from outside the market, there may be universal mechanisms responsible for generating market fluctuations as indicated by the observation of invariant properties. The rigorous demonstration of such a universal law for market behavior is significant for the physics of strongly

interacting complex systems, as it suggests the existence of robust features that are independent of individual details of different systems.

5.2 Distribution of Trading Volume and Number of Trades

Apart from stock price and index fluctuations, one can also measure market activity by looking at the trading volume (the number of shares traded), $V(t)$, and the number of trades, $N(t)$. To obtain the corresponding distributions, we normalize these variables by subtracting the mean and dividing by their standard deviation, such that, $v = \frac{V-\langle V\rangle}{\sqrt{\langle V^2\rangle-\langle V\rangle^2}}$ and $n = \frac{N-\langle N\rangle}{\sqrt{\langle N^2\rangle-\langle N\rangle^2}}$. Figure 5.12 shows the distribution of these two quantities for several stocks, based on daily data for the Bombay Stock Exchange (BSE). As is evident, the distribution is very similar for the different stocks, and the nature of the decay is significantly different from a power law. To better characterize the distribution, we have also looked at the intra-day distributions for volume and number of trades, based on high-frequency data from the NSE. Figure 5.13 shows the distributions of the two quantities for trading conducted on a particular stock in 5 minute intervals. Analysis of data for other stocks show qualitatively similar results. As is clear, both of these distributions are nonmonotonic, and are suggestive of a log-normal form. The fact that these distributions are very similar to each other is not surprising in view of the almost linear relationship between the two (Figure 5.13c). This supports previous observation in major US stock markets that statistical properties of the number of shares traded and the number of trades in a given time interval are closely related [58].

For US markets, power-law tails have been reported for the distribution of both the number of trades [57] and the volume [58]. It has also been claimed that these features are observed on the Paris Bourse, and therefore, these features are as universal as the inverse cubic law for price returns distribution [59, 60]. However, anal-

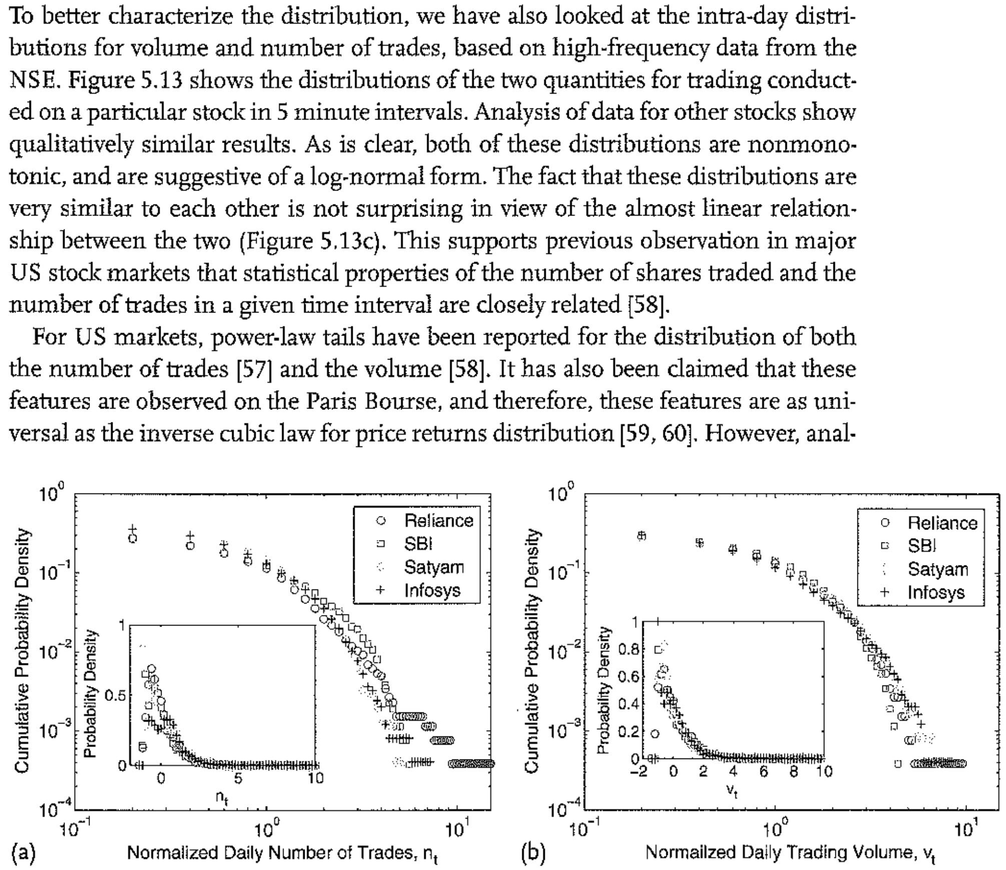


Figure 5.12 Cumulative distribution of the number of trades (top (a)) and the volume of shares traded (top (b)) in a day for four stocks at BSE between July 12, 1995 and January 31, 2006.

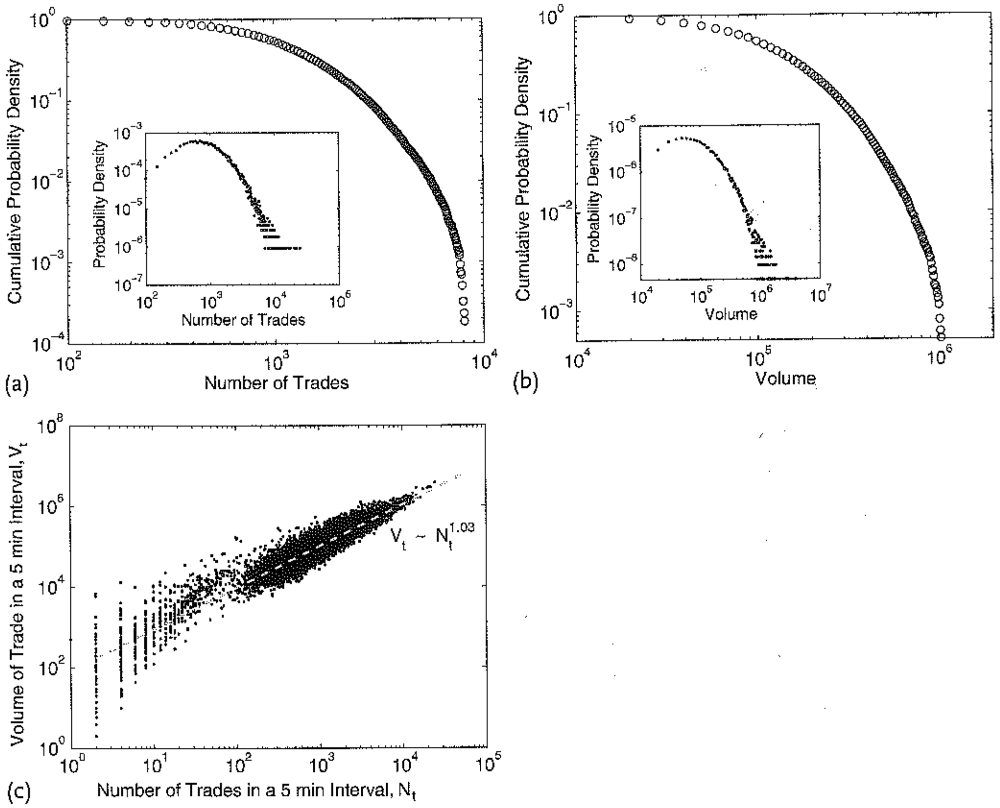


Figure 5.13 Cumulative distribution of the number of trades (a) and the volume of shares traded (b) for a particular stock (Reliance) in 5-minute intervals at NSE between January 1, 2003 to March 31, 2004. (c) shows an almost linear relation between the number of trades in a 5-minute interval and the corresponding trading volume. The broken line indicates the best fit on a doubly logarithmic scale.

ysis of other markets [15, 61] have failed to see any evidence for the universality of the power-law behavior. Our results appear to support the latter assertion that the power-law behavior in this case may not be universal, and the particular form of the distribution of these quantities may be market specific.

5.3 A Model for Reproducing the Power Law Tails of Returns and Activity

There have been several attempts at modeling the dynamics of markets which reproduce at least a few of the above stylized facts. Many of them assume that the price fluctuations are driven by endogenous interactions rather than exogenous factors such as, arrival of news affecting the market and variations in macroeco-

nomic indicators. A widely used approach for such modeling is to consider the market movement to be governed by explicit interactions between agents who are buying and selling assets [18–20, 62, 63]. While this is appealing from the point of view of statistical physics, resembling as it does interactions between spins arranged over a specified network, it is possible that in the market the mediation between agents is done through means of a globally accessible signal, namely the asset price. This is analogous to a mean-field-like simplification of the agent-based model of the market, where each agent is taking decisions based on a common indicator variable. Here, we present a model of market dynamics where the agents do not interact directly with each other, but respond to a global variable defined as price. The price in turn is determined by the relative demand and supply of the underlying asset that is being traded, which is governed by the aggregate behavior of the agents (each of whom can buy, sell or hold at any given time). In the model described here, the trading occurs in a two-step process, with each agent first deciding whether to trade or not at that given instant based on the deviation of the current price from an agent's notion of the "true" price (given by a long-time moving average). This is followed by the agents who have decided to trade choosing to either buy or sell based on the prevalent demand-supply ratio measured by the logarithmic return.

A simplified view of a financial market is that it consists of a large number of agents (say, N) trading in a single asset. During each instant the market is open, a trader may decide to either buy, sell or hold (i.e., remain inactive) based on its information about the market. Thus, considering the time to evolve in discrete units, we can represent the state of each trader i by the variable $S_i(t)$ ($i = 1, \ldots, N$) at a given time instant t. It can take values $+1$, -1 or 0 depending on whether the agent buys or sells *a unit quantity* of asset or decides not to trade at time t, respectively. We assume that the evolution of price in a free market is governed only by the relative supply and demand for the asset. Thus, the price of the asset at any time t, $P(t)$, will rise if the number of agents wishing to buy it (i.e., the demand) exceeds the number wishing to sell it (i.e., supply). Conversely, it will fall when supply outstrips demand. Therefore, the relation between prices at two successive time instants can be expressed as

$$P_{t+1} = \frac{1 + M_t}{1 - M_t} P_t \,, \tag{5.22}$$

where $M_t = \sum_i S_i(t)/N$ is the net demand for the asset, as the state of agents who do not trade is represented by 0 and do not contribute to the sum. This functional form of the time-dependence of price has the following desirable feature: when everyone wants to sell the asset ($M_t = -1$), its price goes to zero, whereas if everyone wants to buy it ($M_t = 1$), the price diverges. When the demand equals supply, the price remains unchanged from its preceding value, indicating an equilibrium situation. The multiplicative form of the function not only ensures that price can never be negative, but also captures the empirical feature of the magnitude of stock price fluctuations in actual markets being proportional to the price. Note that, if the ratio of demand to supply is an uncorrelated stochastic process, price will follow a geometric random walk, as originally suggested by Bachelier [6]. The exact form of the price function (5.22) does not critically affect our results, as we shall discuss later.

Having determined how the price of the asset is determined based on the activity of traders, we now look at how individual agents make their decisions to buy, sell or hold. As mentioned earlier, we do not assume direct interactions between agents, nor do we consider information external to the market to be affecting agent behavior. Thus, the only factor governing the decisions made by the agents at a given time is the asset price (the current value as well as its history up to that time). First, we consider the condition that prompts an agent to trade at a particular time (i.e., $S_i = \pm 1$), rather than hold ($S_i = 0$). The fundamental assumption that we shall use here is that this decision is based on the deviation of the current price at which the asset is being traded from an individual agents notion of the "true" value of the asset. Observation of order book dynamics in markets has shown that the life-time of a limit order is longer, the farther it is from the current bid-ask [64]. In analogy to this we can say that the probability of an agent to trade at a particular price will decrease with the distance of that price from the "true" value of the asset. This notion of the "true" asset price is itself based on information about the price history (as the agents do not have access to any external knowledge related to the value of an asset) and thus can vary with time. The simplest proxy for estimating the "true" value is a long-time moving average of the price time-series, $\langle P_t \rangle_\tau$, with the averaging window size, τ, being a parameter of the model. Our use of the moving average is supported by previous studies that have found the long-time moving average of prices to define an effective potential that is seen to be the determining factor for empirical market dynamics [65].

In light of the above discussion, a simple formulation for the probability of an agent i to trade at time t is

$$P(S_i(t) \neq 0) = \exp\left(-\mu \left|\log \frac{P_t}{\langle P_t \rangle_\tau}\right|\right), \tag{5.23}$$

where μ is a parameter that controls the sensitivity of the agent to the magnitude (i.e., absolute value) of the deviation from the "true" value. This deviation is expressed in terms of a ratio so that, there is no dependence on the scale of measurement. For the limiting case of $\mu = 0$, we get a binary state model, where each agent trades at every instant.

Once an agent decides to trade based on the above dynamics, it has to choose between buying and selling a unit quantity of the asset. We assume that this process is fully dictated by the principle of supply and demand, with agents selling (buying) if there is an excess of demand (supply) resulting in an increase (decrease) of the price in the previous instant. Using the logarithmic return as the measure for price movement, we can use the following simple form for calculating the probability that an agent will sell at a given time t

$$P(S_i(t) = -1) = \frac{1}{1 + \exp\left(-\beta \log \frac{P_t}{P_{t-1}}\right)}. \tag{5.24}$$

The form of the probability function is adopted from that of the Fermi function used in statistical physics, for example for describing the transition probability of spin states in a system at thermal equilibrium. The parameter β, corresponding to "inverse temperature" in the context of Fermi function, is a measure of how strongly the information about price variation influences the decision of a trader. It controls the slope of the function at the transition region where it increases from 0 to 1, with the transition getting sharper as β increases. In the limit $\beta \to \infty$, the function is step-like, such that every trading agent sells (buys) if the price has risen (fallen) in the previous instant. In the other limiting case of $\beta = 0$, the trader buys or sells with equal probability, indicating an insensitivity to the current trend in price movement. The results of the model are robust with respect to variation in β and we shall consider below only the limiting case of $\beta = 0$.

We now report the results of numerical simulations of the model discussed above, reproducing the different stylized facts mentioned earlier. For all runs, the price is assumed to be 1 at the initial time ($t = 0$). The state of every agent is updated at a single time-step or iteration. To obtain the "true" value of the asset at $t = 0$, the simulation is initially run for τ iterations during which the averaging window corresponds to the entire price history. At the end of this step, the actual simulation is started, with the averaging being done over a moving window of fixed size τ.

The variation of the asset price as a result of the model dynamics is shown in Figure 5.14a, which looks qualitatively similar to price (or index) time-series for real markets. The moving average of the price, that is considered to be the notional "true" price for agents in the model, is seen to track a smoothed pattern of price variations, coarse-grained at the time-scale of the averaging window, τ. The price fluctuations, as measured by the normalized logarithmic returns (Figure 5.14b), show large deviations that are significantly greater than that expected from a Gaussian distribution.

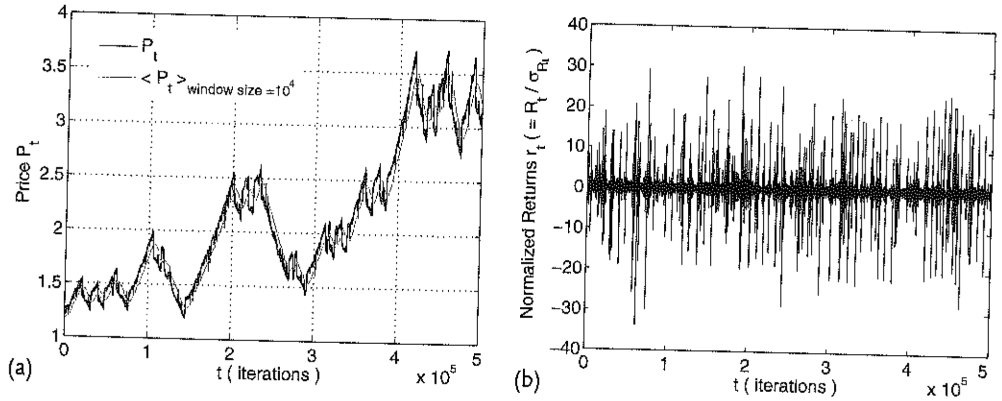


Figure 5.14 (a) Price time-series, along with the moving average of price calculated over a window of size τ, and (b) the corresponding logarithmic returns, normalized by subtracting the mean and dividing by the standard deviation, for a system of $N = 20\,000$ agents. The model is simulated with parameter values $\mu = 100$, $\beta = 0$ and averaging window size, $\tau = 10^4$ iterations.

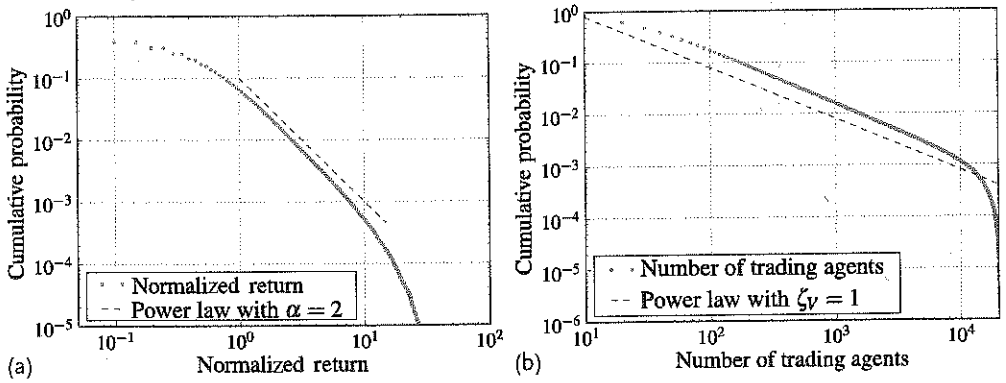


Figure 5.15 Cumulative distributions of (a) normalized returns and (b) trading volume measured in terms of the number of traders at a given time, for a system of $N = 20\,000$ agents. The model is simulated for $T = 200\,000$ iterations, with parameter values $\mu = 100$, $\beta = 0$ and averaging window size, $\tau = 10^4$ iterations. Each distribution is obtained by averaging over 10 realizations.

We now examine the nature of the distribution of price fluctuations by focusing on the cumulative distribution of returns, that is $P(r_t > x)$, shown in Figure 5.15a. We observe that it follows a power law having exponent $\alpha \simeq 2$ over an intermediate range with an exponential cut-off. The quantitative value of the exponent is seen to be unchanged over a large range of variation in the parameter μ and does not depend at all on β. At lower values of μ (viz., $\mu < 10$) the return distribution becomes exponential.

The dynamics leading to an agent choosing whether to trade or not is the crucial component of the model that is necessary for generating the non-Gaussian fluctuation distribution. This can be explicitly shown by considering the special case when $\mu = 0$, where, as already mentioned, the number of traders at any given time is always equal to the total number of agents. Thus, the model is only governed by (5.24), so that the overall dynamics is described by a difference equation or map with a single variable, the net demand (M_t). Analyzing the map, we find that the system exhibits two classes of equilibria, with the transition occurring at the critical value of $\beta = 1$. For $\beta < 1$, the mean value of M is 0, and the price fluctuations follow a Gaussian distribution. When β exceeds 1, the net demand goes to 1 implying that price diverges. This prompts every agent to sell at the next instant, pushing the price to zero, analogous to a market crash. It is a stable equilibrium of the system, corresponding to market failure. This underlines the importance of the dynamics described by (5.23) in reproducing the stylized facts.

As each trader can buy/sell only a unit quantity of asset at a time in the model, the number of trading agents at time t, $V_t = \sum_i |S_i(t)|$, is equivalent to the trading volume at that instant. The cumulative distribution of this variable, shown in Figure 5.15b, has a power-law decay which is terminated by a exponential cut-off due to the finite number of agents in the system. The exponent of the power law, ζ_V, is

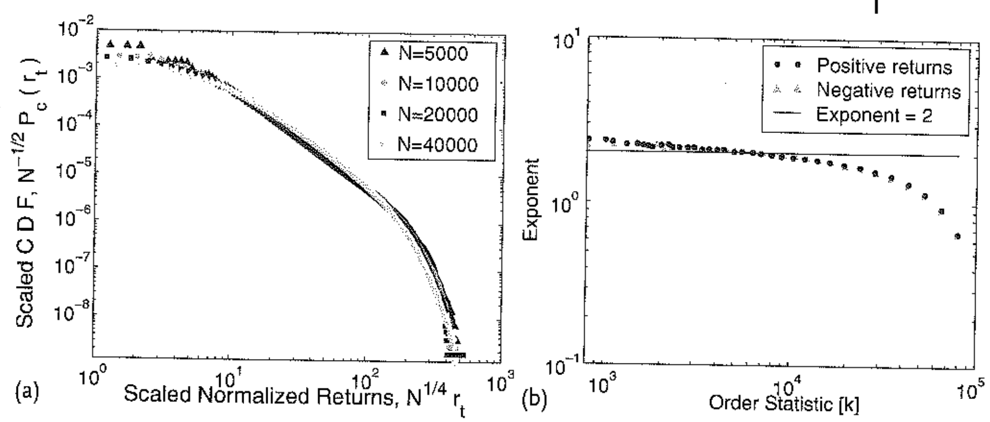


Figure 5.16 (a) Finite size scaling of the distribution of normalized returns for systems varying between $N = 5000$ and $40\,000$ agents, and (b) estimation of the corresponding power-law exponent by the Hill estimator method for a system of $N = 20\,000$ agents. The model is simulated with parameter values $\mu = 100$, $\beta = 0$ and averaging window size, $\tau = 10^4$ iterations.

close to 1, indicating a Zipf's law[3] [1] distribution for the number of agents who are trading at a given instant. As in the case of the return distribution exponent, the quantitative value of the exponent ζ_V is seen to be unchanged over a large range of variation in the parameter μ. The power-law nature of this distribution is more robust, as at lower values of μ (viz., $\mu < 10$), when the return distribution shows exponential behavior, the volume distribution still exhibits a prominent power-law tail.

It is well known that for many systems, their finite size can affect the nature of distributions of the observed variables. In particular, we note that the two distributions considered above have exponential cut-offs that are indicative of the finite number N of agents in the system. In order to explore the role of system size in our results, we perform finite size scaling of the return distribution to verify the robustness of the power-law behavior. This is done by carrying out the simulation at different values of N and trying to see whether the resulting distributions collapse onto a single curve when they are scaled properly. Figure 5.16a shows that for systems between $N = 5000$ and $40\,000$ agents, the returns fall on the same curve, when the abscissa and ordinate are scaled by the system size, raised to an appropriate power. This implies that the power law is not an artifact of finite size systems, but should persist as larger and larger number of agents are considered.

To get a quantitatively accurate estimate of the return exponent, we use the method described by Hill [43]. This involves obtaining the Hill estimator, $\gamma_{k,n}$, from a set of finite samples of a time-series showing power-law distributed fluctuations, with n indicating the number of samples. Using the original time-series $\{x_i\}$, we create a new series by arranging the entries in decreasing order of their magnitude and labeling each entry with the order statistic k, such that the magnitude of x_k is

3) Zipf's law is a special case of the power law where the cumulative probability distribution function $P_c(x) \sim 1/x$.

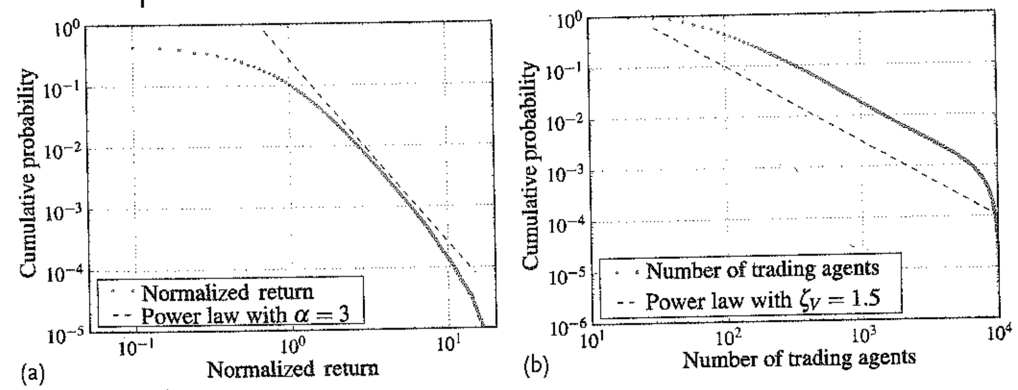


Figure 5.17 Distribution of (a) normalized returns and (b) number of trading agents, for a system of $N = 10\,000$ agents when the parameter μ is randomly selected for each agent from a uniform distribution between [10, 200]. The exponents for the power law seen in both curves agree with the corresponding values seen in actual markets. The model is simulated with parameter values $\beta = 0$ and averaging window size, $\tau = 10^4$ iterations.

larger than that of x_{k+1}. The Hill estimator is calculated as

$$\gamma_{k,n} = \frac{1}{k}\sum_{i}^{k}\log\frac{x_i}{x_{k+1}}\,, \tag{5.25}$$

where $k = 1, \ldots, n-1$. It approaches the inverse of the true value of the power-law exponent as $k \to \infty$ and $\frac{k}{n} \to 0$. Figure 5.16b shows the estimated value of the return distribution exponent, α (i.e., $\gamma_{k,n}^{-1}$) calculated for returns obtained for a system of size $N = 20\,000$ agents. This confirms our previous estimate of $\alpha \simeq 2$ based on least square fitting of the data.

5.3.1
Reproducing the Inverse Cubic Law

So far we have worked in the situation when all the parameter values are constant and uniform for all agents. However, in the real world, agents tend to differ from one another in terms of their response to the same market signal, for example in their decision to trade in a high risk situation. We capture this heterogeneity in agent behavior by using a random distribution of the parameter μ that controls the probability that an agent will trade when the price differs from the "true" value of the asset. A low value of the parameter represents an agent who is relatively indifferent to this deviation. On the other hand, an agent who is extremely sensitive to this difference and refuses to trade when the price goes outside a certain range around the "true" value, is a relatively conservative market player with a higher value of μ.

Figure 5.17 shows the distributions for the return and number of traders when μ for the agents is distributed uniformly over the interval [10, 200] (we have verified that small variations in the bounds of this interval do not change the results). While

the power-law nature is similar to that for the constant parameter case seen earlier, we note that the exponent values are now different and quantitatively match those seen in real markets. In particular, the return distribution reproduces the inverse cubic law, that has been found to be almost universally valid. Surprisingly, we find that the same set of parameters which yield this return exponent, also result in a cumulative distribution for the trading volume (i.e., number of traders) with a power-law exponent $\zeta_V \simeq 1.5$ that is identical to that reported for several markets [58]. Thus, our model suggests that heterogeneity in agent behavior is the key factor behind the observed distributions. It predicts that when the behavior of market players become more homogeneous, as for example, during a market crash event, the return exponent will tend to decrease. Indeed, earlier work [66] has found that during crashes, the exponent for the power-law tail of the distribution of relative prices has a significantly different value from that seen at other times. From the results of our simulations, we predict that for real markets, the return distribution exponent α during a crash will be close to 2, the value obtained in our model when every agent behaves identically.

How sensitive are the above results to the specific forms of the dynamics we have used ? One can test the robustness of these results with respect to the way asset price is defined in the model. We have considered several variations of (5.22), including a quadratic function, viz.,

$$P_{t+1} = \left(\frac{1+M_t}{1-M_t}\right)^2 P_t\,, \tag{5.26}$$

and find the resulting nature of the distributions and the volatility clustering property to be unchanged. The space of parameter values can also be explored for checking the general validity of the results. As already mentioned, the parameter β does not seem to affect the nature, or even the quantitative value of the exponents, of the distributions. The robustness of the results has also been verified with respect to the averaging window size, τ. We find the numerical values of the exponents to be unchanged over the range $\tau = 10^4$–10^6.

It may be pertinent here to discuss the relevance of our observation of an exponential return distribution in the model at lower values of the parameter μ. Although the inverse cubic law is seen to be valid for most markets, it turns out that there are a few cases, such as the Korean market index KOSPI, for which the return distribution is reported to have an exponential form [29]. We suggest that these deviations from the universal behavior can be due to the existence of a high proportion of traders in these markets who are relatively indifferent to large deviations of the price from its "true" value. In other words, the presence of a large number of risk takers in the market can cause the return distribution to have exponentially decaying tails. The fact that for the same set of parameter values, the cumulative distribution of number of traders still shows a power-law decay with exponent 1, prompts us to further predict that, despite deviating from the universal form of the return distribution, the trading volume distribution of these markets will follow a power-law form with ζ_V close to 1.

References

1 Newman, M.E.J. (2005) Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, **46**, 323–351.
2 Clauset, A., Shalizi, C.R., and Newman, M.E.J. (2009) *SIAM Review* **51**(4), 661–703.
3 Mantegna, R.N. and Stanley, H.E. (1999) *Introduction to Econophysics*, Cambridge University Press, Cambridge.
4 Bouchaud, J.P. and Potters, M. (2003) *Theory of Financial Risk and Derivative Pricing*, Second edition, Cambridge University Press, Cambridge.
5 Chatterjee, A. and Chakrabarti, B.K. (eds) (2006) *Econophysics of Stock and other Markets*, Springer, Milan.
6 Bachelier, L. (1900) *Annales Scientifiques de l'École Normale Supérieure Sér*, **3**(17), 21–86.
7 Osborne, M. (1964) The Random Character of Stock Market Prices, chapter *Brownian motion in the stock market*, The MIT Press Cambridge, MA, pp. 100–128.
8 Fama, E.F. (1965) *The Journal of Business*, **38**, 34.
9 Mandelbrot, B. (1963) *The Journal of Business*, **36**(4), 394–419.
10 Mantegna, R.N. and Stanley, H.E. (1995) *Nature*, **376**, 46–49.
11 Gopikrishnan, P., Meyer, M., Nunes Amaral, L.A., and Stanley, H.E. (1998) *Eur. Phys. J. B*, **3**, 139–140.
12 Jansen, D.W. and deVries, C.G. (1991) *The Review of Economics and Statistics*, **73**(1), 18–24.
13 Lux, T. (1996) *Applied Financial Economics*, **6**(6), 463–475.
14 Plerou, V., Gopikrishnan, P., Nunes Amaral, L. A., Meyer, M., and Stanley, H. E. (1999) *Phys. Rev. E*, **60**(6), 6519–6529.
15 Farmer, J.D. and Lillo, F. (2004) *Quantitative Finance*, 4(1), C7–C11.
16 Gopikrishnan, P., Plerou, V., Nunes Amaral, L.A., Meyer, M., and Stanley, H.E. (1999) *Phys. Rev. E*, **60**(5), 5305–5316.
17 Oh, G., Um, C.-J., and Kim, S. (2006) *Physics*, 0601126.
18 Lux, T. and Marchesi, M. (1999) *Nature*, **397**, 498–500.
19 Bornholdt, S. (2001) *Int. J. Mod. Phys. C*, **12**(5), 667–674.
20 Iori, G. (2002) *Journal of Economic Behavior & Organization*, **49**(2), 269–285.
21 Chowdhury, D. and Stauffer, D. (2004) *Eur. Phys. J. B*, **8**(3), 477–482.
22 Sinha, S., Pan, R.K. (2006) How a hit is born: The emergence of popularity from the dynamics of collective choice, (eds A. Chatterjee, A. Chakraborti, B.K. Chakrabarti), *Econophysics and Sociophysics: Trends and Perspectives*, Wiley-VCH, Berlin, p. 417–448.
23 Mandelbrot, B.B. (1997) *Fractals and Scaling in Finance*, Springer, New York.
24 Bacry, E., Delour, J., and Muzy, J.F. (2001) *Phys. Rev. E*, **64**(2), 026103.
25 Bouchaud, J.P. (2005) *Chaos*, **15**, 026104.
26 Storer, R. and Gunner, S.M. (2002) *Int. J. Mod. Phys. C*, **13**(7), 893–897.
27 Wang, B. and Hui, P. (2001) *Eur. Phys. J. B*, **20**(4), 573–579.
28 Kaizoji, T. and Kaizoji, M. (2003) *Advances in Complex Systems*, **6**(3), 303–312.
29 Yang, J.-S., Chae, S., Jung, W.-S., and Moon, H.-T. (2006) *Physica A*, **363**(2), 377–382.
30 Miranda, L.C. and Riera, R. (2001) *Physica A*, **297**(3–4), 509–520.
31 Gleria, I., Matsushita, R., and Silva, S.D. (2002) *Economics Bulletin*, **7**(3), 1–12.
32 Coronel-Brizio, H.F. and Hernandez-Montoya, A.R. (2005) *Revista Mexicana de Fisica*, **51**(1), 27–31.
33 Rak, R., Drozdz, S., and Kwapien, J. (2007) *Physica A*, **374**, 315–324.
34 Jondeau, E. and Rockinger, M. (2003) *Journal of Empirical Finance*, **10**, 559–581.
35 Matia, K., Pal, M., Salunkay, H., and Stanley, H.E. (2004) *Europhys. Lett.*, **66**(6), 909–914.
36 Pan, R.K. and Sinha, S. (2007) *Europhys. Lett.*, **77**, 58004.
37 http://www.nseindia.com/.
38 *Annual Report and Statistics 2005*, World Federation of Exchanges, 2006, p. 77.
39 http://www.nseindia.com/.

40 http://finance.yahoo.com/.
41 Sinha, S., Pan, R.K. (2006) The Power (Law) of Indian Markets: Analysing NSE and BSE trading statistics, (eds A. Chatterjee, B.K. Chakrabarti), *Econophysics of Stock and Other Markets*, Springer, Milan, p. 24–34.
42 Pan, R.K., Sinha, S. (2008) Inverse-cubic law of index fluctuation distribution in Indian markets, *Physica A*, **387**, 2055–2065.
43 Hill, B.M. (1975) *The Annals of Statistics*, **3**(5), 1163–1174.
44 Drees, H., deHaan, L., and Resnick, S. (2000) *The Annals of Statistics*, **28**(1), 254–274.
45 Annual report and statistics 2005 Technical report World Fedration of Exchanges (2006).
46 Pisarenko, V. and Sornette, D. (2006) *Physica A*, **366**, 387–400.
47 Pisarenko, V., Sornette, D., and Rodkin, M. (2004) *Computational Seismology*, **35**, 138–159.
48 Pictet, O.V., Dacorogna, M.M., and Muller, U.A. (1998) *A practical guide to heavy tails chapter Hill, bootstrap and jackknife estimators for heavy tails*, Birkhauser, pp. 283–310.
49 Hall, P. (1990) *Journal of Multivariate Analysis*, **32**, 177–203.
50 Wood, R.A., McInish, T.H., and Ord, J.K. (1985) *The Journal of Finance*, **40**, 723–739.
51 Harris, L. (1986) *Journal of Financial Economics*, **16**, 99–117.
52 Lux, T. (2001) *Applied Financial Economics*, **11**(3), 299–315.
53 Silva, A.C., Prange, R.E., and Yakovenko, V.M. (2004) *Physica A*, **344**, 227–235.
54 Goldstein, M.L., Morris, S.A., and Yen, G.G. (2004) *Eur. Phys. J. B*, 41(2), 255–258.
55 Coronel-Brizio, H. and Hernandez-Montoya, A. (2005) *Physica A*, **354**, 437–449.
56 Pan, R.K. and Sinha, S. (2007) *Phys. Rev. E*, **76**(4), 046116.
57 Plerou, V., Gopikrishnan, P., Amaral, L.A.N., Gabaix, X., Stanley, H.E. (2000) Economic fluctuations and anomalous diffusion. *Phys. Rev. E*, **62**, 3023–3026.
58 Gopikrishnan, P., Plerou, V., Gabaix, X., and Stanley, H.E. (2000) Statistical properties of share volume traded in financial markets, *Phys. Rev. E*, **62**, R4493–R4496.
59 Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H.E. (2003) A theory of power-law distributions in financial market fluctuations, *Nature*, **423**, 267–270.
60 Plerou, V. and Stanley, H.E. (2007) *Phys. Rev. E*, **76**(4), 046109.
61 Racz, E., Eisler, Z. and Kertesz, J.(2009) *Phys. Rev. E*, **79**(6), 068101.
62 Chowdhury, D. and Stauffer, D. (1999) A generalised spin model of financial markets. *Eur. Phys. J. B*, **8**, 477–482.
63 Cont, R. and Bouchaud, J.-P. (2000) Herd behavior and aggregate fluctuations in financial markets. *Macroecon. Dyn.*, 4, 170–196.
64 Potters, M. and Bouchaud, J.-P. (2003) More statistical properties of order books and price impact. *Physica A*, **324**, 133–140.
65 Alfi, V., Coccetti, F., Marotta, M., Pietronero, L., Takayasu, M. (2006) Hidden forces and fluctuations from moving averages: A test study. *Physica A*, **370**, 30–37.
66 Kaizoji T. (2006) A precursor of market crashes: Empirical laws of Japan's internet bubble. *Eur. Phys. J. B*, **50**, 123–127.