An Analysis of Stock Index Distributions of Selected Emerging Markets

Camilleri, Silvio John

University of Malta

2006

This working paper was published as follows:

Camilleri, S.J., 2006, An Analysis of Stock Index Distributions of Selected Emerging Markets, *Bank of Valletta Review*, 33, 33-49, Spring.

An Analysis of Stock Index Distributions of Selected Emerging Markets

Silvio John Camilleri

Banking and Finance Department, FEMA, University of Malta,
Msida, MSD 06, Malta
Email: silvio.j.camilleri@um.edu.mt

February 2006

The author thanks the Buenos Aires Stock Exchange, Cairo and Alexandria Stock Exchange, Bombay Stock Exchange, Jamaica Stock Exchange, National Stock Exchange of Lithuania, Ljubljana Stock Exchange, Malta Stock Exchange, Mauritius Stock Exchange, and Taiwan Stock Exchange Corporation for providing the data. The useful comments from Philip Hans Franses (Erasmus University, Rotterdam) are thankfully acknowledged. Any errors remain the sole responsibility of the author.

An Analysis of Stock Index Distributions of Selected Emerging Markets

Abstract

Stock market data tends to display distinct characteristics commonly known as "stylized facts". These include non-stationarity of price levels, as well as peak-shaped, fat-tailed and heteroskedastic log returns. This paper presents empirical evidence of these characteristics for emerging market indices, spanning over different geographic regions. The results do not disclose asymmetry in the tails of log return distributions in any particular direction. In addition, it is not confirmed that high volatility tends to follow large negative returns.

JEL Classification: G10, G12, G15
Keywords: Emerging Financial Markets, Stylized Facts of Stock Market Data

An Analysis of Stock Index Distributions of Selected Emerging Markets

1. Introduction

The analysis of the return distributions of financial assets is a vital topic in the finance discipline, not only on account of the academic research undertaken in this area, but also due to its relevance for practitioners when making portfolio choices and in risk management processes. When examining stock market price series, the data is typically non-stationary and deviates from the normal distribution. Indeed, stock market data often displays distinct characteristics which are commonly referred to as “stylized facts”.

The aim of this paper is to glean empirical evidence of such characteristics, in respect of various emerging market stock indices. Basic principles of stock market data and relevant literature are reviewed in Section 2. Section 3 includes a description of the methodology, whilst a data description and the limitations of the study are included in Section 4. The empirical results are shown in the subsequent section. Section 6 concludes the analysis.

2. Statistical Principles and Brief Literature Review

Most econometric tools assume stationary characteristics of the data set being analysed. A time series may be defined as stationary if its properties such as the mean and variance are basically unchanged over different sub-samples of the data set. Financial time series tend to deviate from stationarity and they often exhibit a time-changing mean and variance, as outlined by Mills (1999; pp. 37). Researchers usually avoid the application of econometric techniques to non-stationary data given that this could lead to flawed conclusions such as spurious regression results as shown by Granger and Newbold (1974). Therefore non-stationary series are transformed to stationary ones; for example a researcher may difference the series or she may analyse the logarithms of the observed time series. Working with logarithms presents distinct advantages as outlined below:

a) Using logarithms one may model a non-linear relationship, through a linear one. For example the relationship $Y = X^a$, may be transformed to a linear relationship as y = ax, where y ≡ log Y and x ≡ log X;

b) When using linear regression on the logarithmic transformation of the series, the estimated coefficients have an immediate interpretation as elasticities; and

c) When applying a log transformation to the data, the series is “compressed”. For example, a series ranging from 1 to 10,000 will approximately range from 0 to 10 when taking the natural logarithms. This often results in a constant variance for the transformed series – although whether this occurs varies in between cases.

Thus, one approach to modelling a time series of stock prices is to use the continuously compounded return or log return r_t, which in the case of a non-dividend paying asset is equal to:

$$r_t \equiv \log(1+R_t) = \log\left[1+\left(\frac{P_t - P_{t-1}}{P_{t-1}}\right)\right] = \log\frac{P_t}{P_{t-1}} = p_t - p_{t-1} \qquad (1)$$

where P_t is the price level, R_t is the (simple) return level, and $p_t \equiv \log P_t$.

The distribution of log returns often deviates from the normal distribution. Various authors such as Fama (1965) presented empirical evidence which exposed the drawbacks of using a normal distribution to model logarithmic returns. Financial log return distributions tend to be peak-shaped and fat tailed – leptokurtic in statistical terminology. This characteristic of financial log returns has been explained by patterns in the arrival of information, as well as patterns in traders' reactions to news, as discussed in Peters (1991). Dacorogna et. al. (2001; pp 133), in an empirical investigation of USD exchange rate returns, showed that as the frequency of the data increases (say from weekly to hourly) the tails of the distribution become fatter.

Stock markets tend to be characterised by periods of substantial volatility interspersed with other periods of lower volatility. This implies a time-changing variance of returns as reviewed in Bollerslev, Chou, and Kroner (1992). Jacobsen and Dannenburg (2003) used stock market data from various developed countries and showed that this characteristic is not only present in high frequency data, but also in time series of lower frequencies such as monthly data.

Thus, security prices are more likely to follow a martingale process rather than a random walk process. A martingale is a stochastic process which is weaker than a random walk. The random walk model requires uncorrelated price changes, yet a martingale process allows for possible serial dependence in the price movements. The main requirement of a martingale process is that the expected future value of an asset given all available information is the current value. This is summarised as follows:

$$E(P_{t+1} \mid \Phi_t) = P_t \qquad (2)$$

where $E(P_{t+1})$ is the expected price in period t+1 and Φ_t is the information in period t which contains at least the past history of P_t. This results in, at least, semi-strong efficiency, where market prices reflect all publicly available information. A martingale process allows for dependency in higher conditional moments of the price changes, such as the conditional variance. The latter is an empirical feature of financial markets as discussed above.

Franses and van Dijk (2000; pp. 13-19) used stock index data to present empirical evidence of two further characteristics of log returns:

a) Large negative returns are more common than large positive returns. This feature was not confirmed by Longin (1996) in an empirical analysis of US stock market data, and by Jondeau and Rockinger (2003) who studied different stock market indices. The latter authors suggested that the common "perception" that left tails are thicker than the right ones might have been cultivated by the presence of data outliers.

b) Franses and van Dijk (2000) also noted that high volatility often follows large negative returns. The authors also showed that the above two features are not as clearly evident in exchange rate returns data. Further empirical evidence of asymmetric volatility responses in relation to positive and negative returns is found in Koutmos (1999), who used stock price indices from G-7 countries. Yet, DeGennaro and Zhao (1998) found mixed evidence on the relationship between returns and volatility for US stock market data and concluded that this relationship is either "weak or variable".

The aim of this paper is to investigate the degree to which the above characteristics are evident in the index data of selected emerging stock markets.

3. Methodology, Data and Limitations

This study focuses on the degree to which the properties listed below are evident in selected indices. The hypotheses are as follows:

3.1 Original price series are non-stationary, however transforming the series to logarithmic returns induces stationarity;
3.2 Logarithmic returns are not normally distributed – they are peak-shaped and fat tailed;
3.3 Logarithmic returns exhibit a time-changing variance;
3.4 The left tail of the distribution is fatter than the right tail which implies that large negative returns are more common than large positive returns; and
3.5 High volatility often follows large negative returns.

3.1 Stationarity of the Original Series and Logarithmic Returns

One preliminary method through which stationarity of a data set may be inferred is to inquire whether the plot of the data discloses a changing mean and variance for different sub-samples of the series.

The autocorrelation function (ACF) is also related to the stationarity properties of the data, given that a persistently high level of serial correlation is an indication of non-stationarity. The ACF shows the autocorrelations of a data series as a function of a time shift *k*. Thus it measures the extent to which one value of the process is correlated with the previous values. The sample autocorrelation function of the time series x_t, at lag *k* is defined in Mills (1990; pp. 65) as follows:

$$\rho_k = \frac{\sum_{t=k+1}^{n}(x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}, \quad k = 1,2,... \tag{3}$$

The standard errors of the autocorrelation coefficients may be computed using the Bartlett (1946) formula, as quoted in Mills (1990; pp. 65-66):

$$S.E.(\rho_k) = \sqrt{n^{-1}(1 + 2\rho_1^2 + ... + 2\rho_{k-1}^2)} \tag{4}$$

where *n* is the number of observations.

One should note that the above procedures do not constitute formal tests. The standard procedure which is used to infer whether a data set is stationary was discussed by Dickey and Fuller (1979) and is known as the augmented Dickey-Fuller (ADF) test. The first differenced time series is expressed as a function of a constant, (an optional) trend, a lag of the levels, as well as *n* lags of the first difference. Thus:

$$\Delta x_t = f\text{ (constant, trend, } x_{t-1}, \Delta x_{t-1}, \ldots, \Delta x_{t-n}). \tag{5}$$

A test for a unit root may by formulated by comparing the coefficient of x_{t-1} with its standard error. The null hypothesis is that the series contains a unit root and is therefore non-stationary; whilst the alternative hypothesis is that the series does not have a unit root. The critical values which are used in this hypothesis test are those of the augmented Dickey-Fuller statistic.

3.2 Distribution of Logarithmic Returns

The asymmetry of a distribution is measured through the skewness, which is defined as:

$$S\hat{K}_y = \frac{1}{n}\sum_{t=1}^{n}\frac{(r_t - \hat{\mu})^3}{\hat{\sigma}^3} \tag{6}$$

where n is the number of observations, r_t is the log return at time t, μ is the mean, and σ is the standard deviation.

The kurtosis of the distribution indicates whether the data is more peak-shaped or flatter than the normal distribution. The kurtosis is defined as:

$$\hat{K}_y = \frac{1}{n}\sum_{t=1}^{n}\frac{(r_t - \hat{\mu})^4}{\hat{\sigma}^4} \qquad (7)$$

where n is the number of observations, r_t is the log return at time t, μ is the mean, and σ is the standard deviation.

The kurtosis of a normal distribution is equal to 3, and a higher kurtosis value indicates a peak-shaped distribution. Another way in which one may inquire whether a data set is peak-shaped or otherwise, is to look at the location of the central percentile, say, the mid 20% observations starting from the end of the 0.4 percentile to the beginning of 0.6 percentile of the standardised returns. If the mid 20% observations of the data set lie in a narrower range of standardised values as compared to the normal distribution, then the distribution is likely to be peak-shaped.

The final test of normality considered in this analysis is the Jarque-Bera (1980) test. This test jointly considers the skewness and kurtosis of the distribution as follows:

$$JB_y = n\left[\frac{S\hat{K}_y^2}{3!} + \frac{(\hat{K}_y - 3)^2}{4!}\right] \xrightarrow{d} \chi_2^2 \qquad (8)$$

where $S\hat{K}_y$ and $\hat{K}_y$ are the skewness and kurtosis of the distribution respectively, whilst n is the sample size. The statistic is used to test the null hypothesis of a normal distribution, and is χ^2 distributed with two degrees of freedom.

3.3 Heteroskedasticity of Logarithmic returns

The plot of the data might reveal whether the variance tends to change over different sub-samples in the set. Yet, a more formal Lagrange Multiplier (LM) test may also be applied. The data set is regressed on a constant, a lag and an error term as follows:

$$r_t = \beta_0 + \beta_1 r_{t-1} + u_t. \qquad (9)$$

The LM statistic is then used to test whether there are autoregressive conditional heteroskedasticity (ARCH) effects in the above error term u_t, as proposed by Engle (1982). The squared error term u^2_t is auto regressed on q lags as follows:

$$u^2_t = \alpha_0 + \rho_1 u^2_{t-1} + \rho_2 u^2_{t-2} + \ldots + \rho_q u^2_{t-q}. \qquad (10)$$

The null hypothesis of no ARCH effects, i.e. $\rho_1 = \rho_2 = \ldots = \rho_q = 0$, is tested against the alternative hypothesis that $\rho_1 \neq 0, \rho_2 \neq 0, \ldots \rho_q \neq 0$.

3.4 Symmetry of the Tails of the Logarithmic Distribution

The empirical investigation in the next section also inquires whether there is any general trend for a fatter right or left tail as compared to the other one. This is done by comparing the location of the extreme percentiles of the distributions. If these percentiles, say the left 1% and the right 1% of the data lie within approximately the same area, the tails are likely to be symmetric. Yet if one of the percentiles is "squeezed" into a narrower range of standardized values as compared to the other one, then the former tail is likely to be fatter.

3.5 High Volatility often follows Large Negative Returns

The methodology used by Franses and van Dijk (2000) for proving that high volatility tends to follow large negative returns, was to work out the correlation between the squared return at day *t* and the return at day *t-1*. A negative correlation coefficient indicates that the larger returns were preceded by a negative return.[1]

4. Data and Limitations

The data set shows daily closing values of nine emerging markets indices: BOLSA (Argentina), CASE 30 (Egypt), BSE 500 (India), JSE Index (Jamaica), LITIN (Lithuania), SBI 20 (Slovenia), MSE Index (Malta), SEMDEX (Mauritius), and TSEC 50 (Total Return) (Taiwan). The average number of observations for these indices is 1725. The actual number of observations and the starting date for each time series are shown in Table 4.

[1] An alternative methodology to infer the level of asymmetry in volatility was proposed by Engle and Ng (1993). The squared error term of the first order autoregressive process [AR(1)] for log returns series is regressed over a constant and a dummy variable of the lagged sign of the AR(1) error term. The dummy variable takes a value of 1 when the lagged error term of the AR(1) process is negative, whilst it takes a value of zero otherwise. A significantly positive coefficient for the dummy variable is an indicator that high volatility follows negative returns. The results obtained from this methodology were broadly in line with those following the method proposed by Franses and van Dijk (2000).

The data was obtained from the respective exchanges i.e. Buenos Aires Stock Exchange, Cairo and Alexandria Stock Exchange, Bombay Stock Exchange, Jamaica Stock Exchange, National Stock Exchange of Lithuania, Ljubljana Stock Exchange, Malta Stock Exchange, Mauritius Stock Exchange, and Taiwan Stock Exchange Corporation. The indices were compiled by the exchanges, with the exception of TSEC 50 (Total Return) Index which was compiled by FTSE International Limited.

The particular indices were selected in order to achieve a comprehensive cross section of emerging markets within different geographic regions: Africa, Asia, Europe and Latin America. The selection of the actual time span of the data was done in order to minimize non-trading periods and/or missing observations which exceeded five *calendar* days. In those cases where missing observations or non-trading periods exceeding five calendar days remained in the sample, the index did not show any major fluctuation during the particular period. Preliminary plots of the time series did not reveal any outlier observations.

This study is subject to the limitations inherent in analysing security price data including:

a) Stock prices are discrete prices; for example price changes have to be in one-sixteenth of a dollar, or multiples thereof. Possible effects of price discreteness include price clustering. Such effects might still be present to some degree in price series where trading is decimalised, given that in such cases prices still have to be quoted in cents and therefore they are still not continuous. According to Campbell, Lo and MacKinlay (1997; pp. 110-112), the impacts of price discreteness become more evident as the sampling period shortens.

b) When analysing stock market data which spans over long periods of time, one should be aware that the conditions which underlie the pricing process are likely to change. For example, a long sample period is likely to include changes in the composition of stock indices and changes in company structure due to merger and takeover activity. At times changes in the trading procedures and changes in the trading hours might also be present. Dacorogna et. al. (2001; pp.5) referred to these effects as the "breakdown of the permanence hypothesis" and the authors also questioned whether a researcher can actually claim that he is analysing the same market when working with a long time-series. Such effects have to be kept in mind when interpreting the empirical results of market microstructure research.

5. Empirical Results

In this section, the above methodology is used to inquire whether the statistical properties described in Section 2 are present in the data set.

5.1 Stationarity of the Original and Logarithmic Series

The plots of the original price series give a preliminary indication that the series are not stationary due to a time changing mean and / or variance. The time series plots for CASE 30 and JSE index are shown in Figures 1 and 2 as examples – the other index plots are not being reproduced for the sake of brevity.

Figure 1: CASE 30 Index

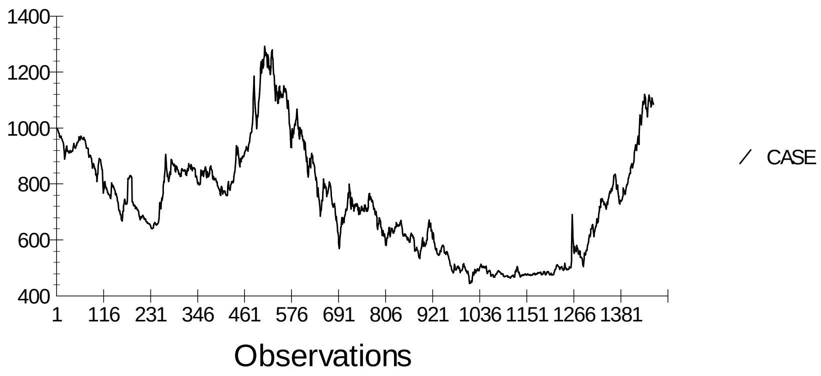
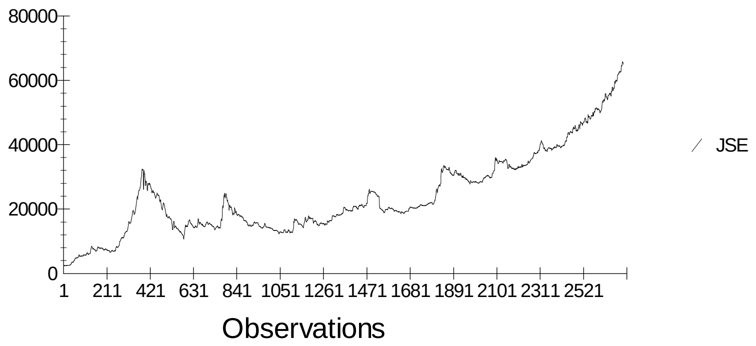


Figure 2: JSE Index



Another indication of non-stationarity is the slowly declining autocorrelation coefficients. Table 1 shows the autocorrelation coefficients and standard errors for the five lags, lag 20 and lag 30 of each index. In all cases, the coefficients remain significant (at the 5% level) till lag 30, indicating that it is not advisable to analyse the original price levels due to non-stationarity.

As noted in Section 2, log return characteristics are often more suitable for the application of econometric techniques given that they tend to be closer to stationarity. The plots of the log returns of BOLSA and TSEC 50 are shown in Figures 3 and 4 – the other plots are not being shown for the sake of conciseness. The plots visually demonstrate that log returns have a mean of approximately zero, or perhaps slightly positive. In addition, most of the plots such as BOLSA disclose a time changing variance, where periods of a relatively low variance alternated with others of higher variance. The time series plot which was visually closest to a constant variance is TSEC 50, yet, even in this case a time-changing variance is plausible.

The autocorrelation coefficients and standard errors for various lags of the log returns of the indices are shown in Table 2. Taking log returns of the time series reduced the level of the serial correlation, even if some of the coefficients remained significant. Indeed, when inspecting the autocorrelation coefficients for the first 30 lags, JSE, MSE, and SEMDEX had at least 10 coefficients which were significant at the 5% level.

Table 1: Autocorrelation Coefficients and Standard Errors of the Original Levels

Order	Autocorrelation Coefficient	Standard Error	Autocorrelation Coefficient	Standard Error	Autocorrelation Coefficient	Standard Error
	BOLSA		**CASE 30**		**BSE 500**	
1	0.994	0.019	0.996	0.026	0.994	0.029
2	0.988	0.033	0.991	0.045	0.988	0.050
3	0.982	0.042	0.986	0.058	0.981	0.064
4	0.975	0.049	0.981	0.069	0.974	0.075
5	0.969	0.056	0.976	0.078	0.967	0.085
20	0.859	0.110	0.906	0.156	0.844	0.167
30	0.787	0.130	0.860	0.187	0.763	0.197
	JSE Index		**LITIN**		**SBI 20**	
1	0.997	0.019	0.994	0.045	0.998	0.019
2	0.994	0.033	0.988	0.078	0.995	0.033
3	0.991	0.043	0.982	0.101	0.992	0.043
4	0.989	0.051	0.975	0.119	0.990	0.051
5	0.986	0.057	0.968	0.134	0.987	0.058
20	0.943	0.117	0.854	0.264	0.945	0.117
30	0.913	0.141	0.769	0.312	0.918	0.142
	MSE Index		**SEMDEX**		**TSEC 50**	
1	0.999	0.029	0.996	0.026	0.995	0.026
2	0.997	0.051	0.991	0.045	0.990	0.045
3	0.995	0.065	0.985	0.057	0.985	0.057
4	0.993	0.077	0.980	0.068	0.980	0.068
5	0.991	0.087	0.974	0.077	0.975	0.077
20	0.955	0.179	0.879	0.153	0.896	0.153
30	0.924	0.217	0.812	0.181	0.841	0.183

Figure 3: BOLSA Log Returns

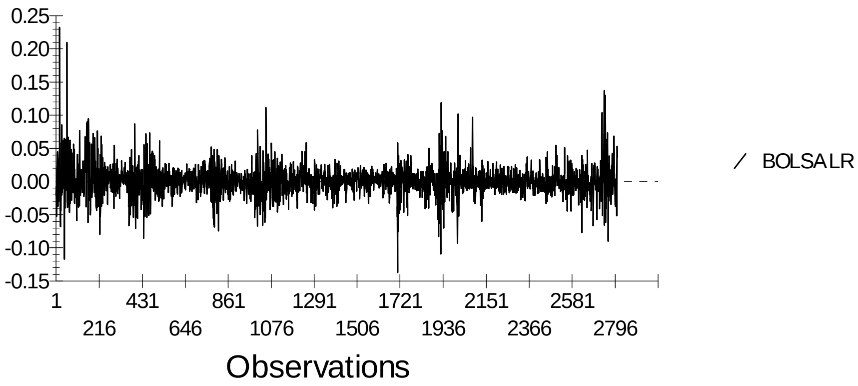
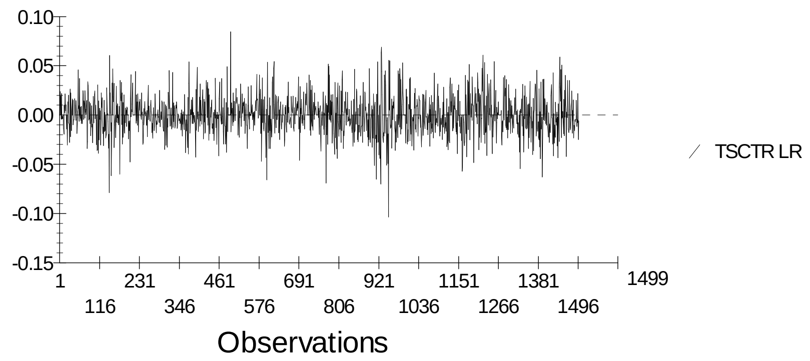


Figure 4: TSEC 50 Log Returns



Therefore, the autocorrelation test confirms that log returns are better candidates for analysis purposes than the original series. Yet, given that as discussed above, these procedures are not formal tests, Augmented Dickey-Fuller tests were used as an indication of the stationarity (or otherwise) of the log returns.

In applying this testing procedure to the data, specifications without a trend were selected, given that the plots of the log returns suggest that it is unlikely that these series include a trend. The ADF results are shown in Table 3. The values of the test statistic as compared to the 95% critical value of the augmented Dickey-Fuller statistic permit rejection of the null hypothesis of a unit root for all the nine indices. This indicates that the log returns series are difference stationary. Overall, the above tests indicate that it is reasonable to analyse the log returns series.

Table 2: Autocorrelation Coefficients and Standard Errors of Log Returns

Order	Autocorrelation Coefficient	Standard Error	Autocorrelation Coefficient	Standard Error	Autocorrelation Coefficient	Standard Error
	BOLSA		**CASE 30**		**BSE 500**	
1	0.173 *	0.018	0.214 *	0.026	0.109 *	0.028
2	-0.019	0.019	-0.020	0.027	0.028	0.029
3	0.029	0.019	0.009	0.027	0.015	0.029
8	0.020	0.019	-0.013	0.027	0.024	0.029
9	0.038	0.019	-0.011	0.027	0.099 *	0.029
10	0.065 *	0.019	0.015	0.027	0.069 *	0.029
15	0.019	0.019	-0.006	0.027	0.000	0.029
20	0.030	0.019	-0.016	0.027	-0.064	0.029
30	-0.003	0.019	0.070 *	0.028	-0.020	0.030
	JSE Index		**LITIN**		**SBI 20**	
1	0.475 *	0.019	0.039	0.045	0.170 *	0.019
2	0.200 *	0.023	0.121 *	0.045	0.054 *	0.019
3	0.057 *	0.023	0.023	0.046	0.010	0.019
8	0.043	0.023	0.039	0.046	0.053 *	0.020
9	0.064 *	0.023	0.001	0.046	0.026	0.020
10	0.054 *	0.023	-0.021	0.046	0.070 *	0.020
15	0.033	0.024	0.032	0.046	0.027	0.020
20	0.085 *	0.024	-0.016	0.047	0.025	0.020
30	0.015	0.024	0.001	0.048	-0.007	0.020
	MSE Index		**SEMDEX**		**TSEC 50**	
1	0.399 *	0.029	0.358 *	0.025	0.029	0.025
2	0.146 *	0.033	0.184 *	0.028	0.058 *	0.025
3	0.009	0.034	0.126 *	0.029	0.031	0.025
8	0.085 *	0.034	0.061 *	0.030	0.008	0.026
9	0.073 *	0.034	0.055	0.030	0.004	0.026
10	0.125 *	0.034	0.044	0.030	0.026	0.026
15	0.061	0.036	0.076 *	0.030	0.056 *	0.026
20	0.084 *	0.036	0.021	0.031	0.025	0.026
30	0.074	0.037	-0.018	0.031	0.015	0.026

(*) indicates significance at the 5% level.

Table 3: Augmented Dickey-Fuller Test Results

Index (Log Returns)	Order Selection (AICC)	Test Statistic	95% Critical Value
BOLSA	2	-28.345	-2.863
CASE 30	1	-25.603	-2.864
BSE 500	0	-31.012	-2.864
JSE Index	2	-25.164	-2.863
LITIN	2	-11.237	-2.868
SBI 20	5	-18.063	-2.863
MSE Index	0	-22.281	-2.865
SEMDEX	3	-16.568	-2.864
TSEC 50	4	-17.065	-2.864

5.2 Distribution of Logarithmic Returns

The histograms of the log returns of the indices indicate that the data is peak-shaped and perhaps fat-tailed. Only those two histograms which most prominently displayed these characteristics are being reproduced for the sake of conciseness. These are shown in Figures 5 and 6. Normal distributions are superimposed on the histograms for ease of comparison.

Figure 5:

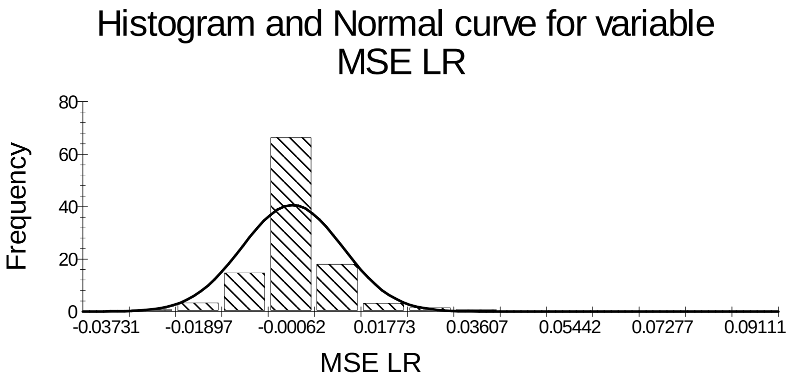


Figure 6:

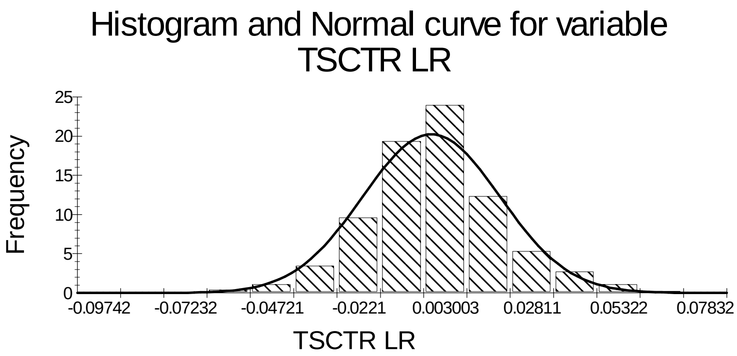


Table 4 shows the basic characteristics of the log return distributions of all nine indices.

Table 4: Basic Characteristics of Log Returns

Country	Argentina	Egypt	India	Jamaica	Lithuania	Slovenia	Malta	Mauritius	Taiwan
Index	BOLSA	CASE 30	BSE 500	JSE Index	LITIN	SBI 20	MSE Index	SEMDEX	TSEC 50
Initial Observation	2-Jan-91	1-Jan-98	1-Feb-99	2-Jan-91	2-Jan-02	7-Jan-93	18-May-98	25-Nov-97	1-Apr-97
No. of Observations	2807	1460	1207	2712	485	2686	1166	1501	1498
Mean	0.0008	0.0001	0.0005	0.0012	0.0007	0.0006	0.0005	0.0002	-0.0002
Standard Deviation	0.0231	0.0168	0.0171	0.0137	0.0119	0.0141	0.0098	0.0046	0.0197
Skewness	0.9669	0.8811	-0.4209	0.9422	-0.1715	0.4201	2.6681	0.7222	0.0697
Excess Kurtosis	9.8529	13.4197	2.0853	12.9245	5.1688	21.3521	20.2589	10.6578	1.4655
Jarque-Bera Statistic	11791.68	11144.28	254.32	19277.08	542.28	51103.24	21323.17	7234.54	135.26
Minimum	-0.1367	-0.1098	-0.0735	-0.1245	-0.0738	-0.1161	-0.0419	-0.0282	-0.1037
Maximum	0.2321	0.1837	0.0693	0.1085	0.0585	0.1893	0.0957	0.0382	0.0846
Coeff. Of Variation	28.4726	302.9185	32.2536	11.4520	16.0366	22.1661	19.9678	20.3734	95.3220
0.01Percentile (of standardised returns)	-2.5069	-2.2938	-2.8627	-2.5989	-2.5461	-2.6888	-2.3255	-2.7730	-2.4954
0.1 Percentile	-1.0356	-1.0433	-1.1746	-0.8620	-1.0046	-0.8872	-0.8547	-0.9787	-1.1508
0.4 Percentile	-0.1780	-0.2101	-0.1246	-0.1750	-0.1879	-0.1371	-0.1519	-0.1788	-0.2136
0.6 Percentile	0.1381	0.1215	0.2441	0.0296	0.1167	0.1176	0.0416	0.1022	0.1180
0.9 Percentile	0.9740	1.1324	1.1220	0.9643	1.0741	0.9067	0.7699	0.9670	1.2478
0.99 Percentile	3.0915	2.4371	2.6199	3.4539	2.5713	2.9567	3.5933	3.0423	2.6342
LM(1) Test for Heteroskedasticity	212.8591	291.2590	126.0123	161.9786	11.9594	314.7868	71.8324	119.5308	25.8294
Correlation (r^2_t, r_{t-1})	0.1174	0.2003	-0.2101	0.1657	-0.0049	-0.1216	0.2313	0.1076	-0.0434

The excess kurtosis values show that in all cases, log return distributions are peak-shaped. An alternative way in which one may infer whether a distribution is peak-shaped is by looking at the location of the standardised values of the mid-observations. The mid-20% observations in a normal distribution lie between the standardised values of +/- 0.251. In comparison, the mid-20% (standardised) observations for all indices lie within a narrower range of standardised values, as shown in Table 4. This is an alternative indication that these distributions are somewhat peak-shaped.

In inquiring whether the distributions of the index log returns are fat-tailed, the location of the "extreme percentiles" may be compared to that of the normal distribution. The 0.01 and 0.99 percentiles in a normal distribution occur at -/+ 2.326. With the exception of the CASE 30 index, the 0.01 percentile occurs "earlier" than expected, indicating fat left tails for eight of the indices being analysed. The location of the 0.99 percentile for the distributions of all indices indicates that the right tail is fatter than the normal one, given that we enter this percentile "later" than expected. This confirms that the tails of the log returns are fatter than normal. Yet, the distributions become narrower than the normal distribution as we move further towards the centre, say when the location of the 0.1 and 0.9 percentiles is considered. The distributions then become "fat" again in the centre, given that they are peak-shaped as discussed above.

Seven of the indices are positively skewed, whilst BSE 500 and LITIN have a negative skewness. Finally, the Jarque-Bera statistics are large enough to enable the rejection of the null hypothesis of normality for all the distributions, at the 99% level of confidence.

5.3 Heteroskedasticity of Logarithmic Returns

As noted above, the plots of logarithmic returns for the indices show that it is quite plausible that both series feature a time-changing variance.

The LM statistics for the indices are shown in Table 4 through an order 1 test. These statistics are compared to the 95% critical value of the χ^2 distribution at the respective degrees of freedom. The values of the LM statistics are high enough to permit rejection of the null hypothesis of no ARCH effects. This is a sign of heteroskedastic time series where large returns tend to occur in clusters.

5.4 Symmetry of the Tails of the Logarithmic Distribution

The histograms of the indices did not visually indicate that the left tail is fatter than the right one, as suggested by Franses and van Dijk (2000). Indeed, comparing the location of the 0.01 percentile with that of the 0.99 percentile (which should be equidistant from zero in a symmetric distribution) indicates a fatter right tail for all indices except BSE 500. When comparing the location of the 0.1 and the 0.9 percentiles, the evidence in favour of fatter right tails declines, given that BOLSA, BSE, MSE and SEMDEX indicate a fatter left tail. Therefore, the empirical results for these indices are in line with the suggestions of Longin (1996) and Jondeau and Rockinger (2003), that asymmetry in the tails in some particular direction is not a general characteristic of stock market returns.

5.5 High Volatility tends to follow Large Negative Returns

Table 4 also reports the correlation coefficients for the squared return r^2_t with the lagged return r_{t-1}. The correlation is negative only in case of four of the nine indices being analysed, and overall this does not confirm the observations of Koutmos (1999) and Franses and van Dijk (2000) that high volatility often follows large negative returns[2]. This is however in line with the conclusion of DeGennaro and Zhao (1998).

Thus, the result that high volatility does not seem to follow large negative returns in emerging markets may be due to no relationship between these variables in the first place, or because high volatility tends to be a more common feature in emerging stock markets, and therefore it tends to follow both negative and positive large returns. The latter hypothesis may be explained by the notion that stock market volatility tends to be interconnected with macroeconomic volatility[3].

Another particular feature of the results obtained with respect to this characteristic is that if the countries are grouped by geographic regions, some patterns emerge. For instance African countries (Egypt and Mauritius) show a positive correlation between volatility and lagged returns. The same applies for Argentina and Jamaica. Asian countries (India and Taiwan) reveal a negative correlation between volatility and lagged returns. The results are somewhat mixed in case of European countries where this relationship is positive in case of Malta, negative for Slovenia and (slightly) negative for Lithuania. Factors which contribute to these differences may include the shorter sample period for Lithuania, and the relatively low market activity on the Malta Stock Exchange as compared to the other exchanges. Overall, sub-dividing the countries in geographic regions reveals similar trends, which may be taken as an indication of the interdependence of proximate markets, although this entails a more rigorous investigation

[2] The correlation r^2_t, r_{t-3} was also estimated and the signs of the coefficients were unchanged. When the correlation r^2_t, r_{t-5} was considered, the signs of the coefficients were confirmed again, with the exception of SBI 20 and SEMDEX.

[3] For instance Morelli (2002) used UK data to present empirical evidence of the interrelationship between the conditional volatility of the stock market and that for various macroeconomic variables.

given that the sample periods at hand differ across the indices.

6. Conclusion

This paper has presented empirical evidence of the "stylized facts" of stock market data, by using selected emerging market indices. A brief exposition of the characteristics and the relevant literature relating to stock market time series was presented in Section 2. The methodology, data set and limitations were subsequently discussed. The empirical results confirmed that stock price levels are often non-stationary and that it is more reasonable to apply econometric tools to the log returns. It was also confirmed that the latter tend to be peak-shaped, fat-tailed and heteroskedastic. The empirical results, did not confirm the observations of other authors regarding the asymmetry of the distribution tails in some particular direction and that high volatility tends to follow large negative returns. Yet similar patterns for the latter characteristic were found over different geographic regions.

In interpreting the above results, one should keep in mind that they might be sensitive to differing sampling intervals, as found for instance by Balaban, Ouenniche and Politou (2005). The use of higher frequency data rather than daily price series might result in even more pronounced deviations from normality.

The deviations of stock return distributions from normality are relevant for portfolio selection and risk management decisions. In modelling the price risk of financial assets, particular attention should be devoted to the tails of the distributions since these constitute the largest price fluctuations and are thus highly relevant for the risk management function. The modelling of these extreme fluctuations may require more focused econometric models, and this issue provides an interesting avenue for further research.

References

Balaban, E., Ouenniche, J., and Politou, D. (2005) "A Note On Return Distribution Of UK Stock Indices," *Applied Economics Letters*, 12, 573-576.

Bartlett, M.S., 1946, On the Theoretical Specification and Sampling Properties of Autocorrelated Time Series, *Journal of the Royal Statistical Society*, Series B, 8, 27-41.

Bollerslev, T., R.Y. Chou, and K.F. Kroner, 1992, ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence, *Journal of Econometrics*, 52, 5-59.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton, NJ: Princeton University Press.

Dacorogna, M., R. Gençay, U. Müller, R. Olsen, O. Pictet, 2001, *An Introduction to High-Frequency Finance*, San Diego CA, Academic Press.

DeGennaro, R.P., and Y.L. Zhao, 1998, Stock Returns and Volatility: Another Look, *Journal of Economics and Finance*, 22(1), 5-18, Spring.

Dickey, D.A., and W.A. Fuller, 1979, Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, 74, 427-431.

Engle, R.F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation, *Econometrica*, 50, 987-1008.

Engle, R.F., and Ng, V.K. (1993) “Measuring and Testing the Impact of News on Volatility”, *The Journal of Finance*, 48(5), 1749-1778.

Fama, E.F, 1965, The Behavior of Stock Market Prices, *Journal of Business*, 38, 34-105.

Franses, P.H., and D. van Dijk, 2000, *Non-Linear Time Series Models in Empirical Finance*, Cambridge and New York: Cambridge University Press.

Granger C.W.J., and P. Newbold, 1974, Spurious Regressions in Econometrics, *Journal of Econometrics*, 2, 111-120.

Jacobsen, B., and D. Dannenburg, 2003, Volatility Clustering in Monthly Stock Returns, *Journal of Empirical Finance*, 10, 479-503.

Jarque, C.M., and A.K. Bera, 1980, Efficient Tests for Normality, Homoscedasticity, and Serial Independence of Regression Residuals, *Economics Letters*, 6, 255-259.

Jondeau, E., and M. Rockinger, 2003, Testing for Differences in the Tails of Stock-Market Returns, *Journal of Empirical Finance*, 10, 559-581.

Koutmos, G., 1999, Asymmetric Index Stock Returns: Evidence from the G-7, *Applied Economics Letters*, 6, 817-820.

Longin, F.M., 1996, The Asymptotic Distribution Of Extreme Stock Market Returns, *Journal of Business*, 69, 383– 408.

Mills, T.C., 1990, *Time Series Techniques for Economists*, Cambridge and New York: Cambridge University Press.

Mills, T.C., 1999, *The Econometric Modelling of Financial Time Series*, Cambridge and New York: Cambridge University Press.

Morelli, D., 2002, The Relationship Between Conditional Stock Market Volatility and Conditional Macroeconomic Volatility. Empirical Evidence Based on UK Data, *International Review of Financial Analysis*, 11, 101-110.

Peters, E., 1991, Chaos and Order in the Capital Markets. *A New View of Cycles, Prices, and Market Volatility*, New York: John Wiley and Sons.