The Structure of International Stock Market Returns

João A. Bastos and Jorge Caiado*
Centre for Applied Mathematics and Economics
School of Economics and Management
Technical University of Lisbon
Rua do Quelhas 6, 1200-781 Lisbon (Portugal)

Abstract

The behavior of international stock market returns in terms of rate of return, unconditional volatility, skewness, excess kurtosis, serial dependence and long-memory is examined. A factor analysis approach is employed to identify the underlying dimensions of stock market returns. In our approach, the factors are estimated not from the observed historical returns but from their empirical properties, without imposing any restriction about the time dependence of the observations. To identify clusters of markets and multivariate outliers, factor analysis is then used to generate factor scores. The findings suggest the existence of meaningful factors which determine the differences in terms of the dependence structure between developed and emerging market returns.

Keywords: Developed and emerging stock markets; Empirical properties of returns, Factor analysis; Serial depedence; Long-memory.

JEL classification: C13; G15.

1 Introduction

International stock return comovements has become an important research area in international finance for several reasons. Investors are interested in international stock market relationships for portfolio diversification and risk management purposes. Economists and finance analysts are interested in these relationships to investigate the comovement structure of countries and to identify groups of countries with similar comovement characteristics as a result of increasing market integration. Comovements of returns and volatility in international stock markets

* *E-mail correspondence address:* jcaiado@iseg.utl.pt

has also motivated financial researchers to develop statistical methods to study the behavior of returns and to identify the sources of return covariation.

During the last years, several statistical methods have been used to investigate the comovements and linkages among stock markets. These include correlation methods (Lin et al., 1994, Longin and Solnik, 1995, Karolyi and Stulz, 1996, Morana and Beltratti, 2008), vector error correction and cointegration analysis (Bessler and Yang, 2003, Syriopoulos, 2004, Tahai et al., 2004, Voronkova, 2004, Rita and Costantini, 2006), factor models (Engle and Susmel, 1993, King et al., 1994, Hui, 2005, Bekaert et al., 2010) and cluster analysis (Bonnano et al., 2004, Caiado and Crato, 2007).

Motivated by these issues, we investigate the common pattern of a broad range of developed markets and emerging markets using a statistical factor analysis approach. Due to the complexity of multivariate parametric models, statistical factor analysis is one of the widely used dimension reduction methods to capture common dynamic features in multiple asset returns. In the traditional statistical factor analysis (Connor and Korajczyk, 1988, Chan et al., 1998, Tsay, 2005), the factors are extracted from the covariance or correlation matrix of the historical returns assuming that the data have no serial correlation. This assumption is often violated in high-frequency financial asset returns. To avoid this problem, some researchers suggest the use of a parametric model (such as ARMA, VAR or VARMA model) to remove the time dependency of the observations and apply the factor analysis to the residual series. However, as pointed out by Tsay (2005) among others, the correlations of the residual series are often very close to the correlations of the original data, and therefore this procedure may be redundant.

Our study differs from previous work in two ways. First, we estimate the latent or unobserved factors not from the observed returns but from their empirical properties such as mean, standard deviation, skewness, kurtosis, linear and nonlinear dependence, and long-memory. In this case, one may use stock returns or market index returns with high frequency without imposing any restriction about the dynamic dependence of observations on factor analysis. Second, the factor loadings, which in our study represent the correlation of the stock return properties with the derived factors, are used to compute the factor scores for each of the stock markets under consideration. These factor scores are then used in subsequent analyses to identify clusters of countries and multivariate outliers. This procedure allows to describe the dynamic structure of multiple returns in terms of a few factors.

The outline of the article is as follows. Section 2 presents the inputs to the factor analysis: the empirical properties of stock returns. Section 3 provides a description of the statistical methodology used. Section 4 describes the data and explores the univariate statistics. Section 5 presents the empirical findings on the factor analysis technique. Section 6 summarizes and concludes.

2 The inputs to the factor analysis

It is well known that financial time series exhibit stylized facts and statistical features. We describe various empirical properties of stock returns such as distributional properties, short-term dependence and long-memory behavior. These features are then used as inputs to the factor analysis.

2.1 Distributional properties of returns

Let P_t denote the price of an asset at time t. The continuously compounded return (or log return) from time $t-1$ to t is defined as $r_t = \ln(P_t/P_{t-1})$. Standard univariate descriptive statistics of asset returns include the mean, the standard deviation, the skewness and the excess kurtosis of returns.

Denote by n the number of observed returns, the mean is computed as the average log return,

$$\overline{r} = \frac{1}{n}\sum_{t=1}^{n} r_t. \tag{1}$$

The standard deviation, or unconditional volatility, is a measure of dispersion in the return series and is usually considered as a proxy of asset risk. The sample standard deviation is given by

$$s = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}\left(r_t - \overline{r}\right)^2}. \tag{2}$$

The skewness is the coefficient of asymmetry of the distribution of the return series. The sample skewness is given by

$$S = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{r_t - \overline{r}}{\widehat{\sigma}}\right)^3, \tag{3}$$

where $\widehat{\sigma} = s\sqrt{(n-1)/n}$. The kurtosis measures the "fatness" of the tails of the returns distribution. The sample kurtosis is computed as

$$K = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{r_t - \overline{r}}{\widehat{\sigma}}\right)^4. \tag{4}$$

The quantity $K-3$ is sometimes called the excess kurtosis. If the data are normally distributed, the skewness and excess kurtosis should be close to zero. A distribution with positive excess kurtosis has heavy tails, whereas a distribution with negative excess kurtosis has short tails. In many empirical studies, the distribution of log returns usually has fatter tails than the normal distribution, which means that extreme events occur more often than would be predicted from a normal distribution. For instance, it is well known that emerging market returns depart from the normal distribution (Harvey, 1995 and Bekaert and Harvey, 1997). For a more detailed discussion of the distributional properties of returns, see, for instance, Cont (2001).

2.2 Short-term dependence

The short-term serial dependence (also known as short-term serial correlation) describes the low-order correlation structure of a time series. In our study, the presence of short-term linear dependence in the stock markets is examined by the autocorrelations of the return series. For financial data, the autocorrelations of returns are zero or very close to zero, which is consistent with the random walk or martingale hypothesis. However, returns to equity indices often do exhibit some serial correlation (Lo and MacKinlay, 1988, Poterba and Summers, 1988, Campbell et al., 1996). Also the random walk hypothesis is more frequently violated in emerging markets than in developed markets (see, e.g., Bastos and Caiado, 2009).

The presence of nonlinear dependence and apparent autoregressive heteroskedasticity effects (Engle, 1982) is judged by the autocorrelations of the squared returns or absolute returns. In contrast to the autocorrelations in returns, which are typically not significant, the autocorrelations for the squared returns or absolute returns are generally positive and significative for a substantial number of lags. This stylized fact is also known as *volatility clustering*, which means that large (small) volatility is often followed by large (small) volatility. In addition, the autocorrelations in the absolute returns are generally higher than the autocorrelations in squared returns, especially for stock market indices (Frances and van Dijk, 2000).

The hypothesis of no autocorrelation up to order m in the returns (absolute returns) is tested using the Ljung-Box modified $Q(m)$-statistic:

$$Q(m) = n(n+2)\sum_{s=1}^{m} \frac{\widehat{\rho}_s^2}{n-s}, \tag{5}$$

where $\widehat{\rho}_s$ denotes the sample autocorrelation of the returns (absolute returns) at lag s. The choice of $m \approx \ln(n)$ may be appropriate for better power properties (Tsay, 2005).

2.3 Long-memory

Many time series exhibit long-memory or long-range dependence behavior (Beran, 1994). More formally, a stationary process x_t exhibits long-memory with memory parameter d if its spectral density function $f(\omega)$ satisfies

$$f(\omega) \sim C\omega^{-2d}, \text{ as } \omega \to 0, \tag{6}$$

where C is a positive finite constant and ω denotes the frequency. When $d < 0.5$ its autocorrelation function ρ_k decays at a hyperbolic rate, i.e.

$$\rho_k \sim C_\rho k^{2d-1}, \tag{7}$$

where C_ρ is a constant with respect to k. If $0 < d < 0.5$, the process has long memory. If $d = 0$, the process has no memory. If $-0.5 < d < 0$, the process has intermediate memory. For $d > 0.5$, the process is no longer covariance stationary.

Geweke and Porter-Hudak (1983) proposed a semi-parametric method to estimate the long-memory parameter d. Under condition (6), the statistical method consists in estimating d using a log-periodogram regression,

$$\ln I(\omega_j) = a - b\ln\left(4\sin^2(\omega_j/2)\right) + \varepsilon_j, \;\; j = 1, ..., l, \tag{8}$$

where $I(\omega_j) = (2\pi n)^{-1}\left|\sum_{t=1}^{n} x_t e^{it\omega}\right|^2$ is the periodogram at harmonic frequency $\omega_j = 2\pi j/n$, $\varepsilon_j = \ln(I(\omega_j)/f(\omega_j))$, and l is the number of low-frequency ordinates used for the regression. Geweke and Porter-Hudak (1983) showed that for $l = n^\lambda$, $0 < \lambda < 1$, the least squares estimate of b of the regression (8) provides a consistent estimate of d, and the usual t-statistic can be employed to test the null hypothesis of no long-memory. For a more detailed discussion on long memory processes, see, for instance, Baillie (1996).

Of particular interest in economics and finance is the long memory behavior of absolute stock returns and squared returns. Many empirical studies have noticed an apparent stylized fact of the very slowly decaying autocorrelations for absolute (or squared) returns. As noted by Ding et al. (1993) and Granger and Ding (1996), the evidence of long memory is stronger for $|r_t|$ than for r_t^2. Using price series from various stock markets and commodity prices, Granger and Ding (1996) showed that $|r_t|$ have the properties of an $I(d)$ process with d values around 0.45.

Some studies have investigated the presence of long range dependence in daily, weekly and monthly international equity market indices. The empirical findings are mixed. Crato (1994) studied the existence of long memory in stock market indices of the G-7 countries. The results showed evidence of long-memory only in one market. Using heteroskedastic-robust testing methods, the modified rescaled range analysis and the variance ratio test, Cow et al. (1996) found no evidence of long-term memory in 22 international equity return indexes. Sadique and Silvapulle (2001) found evidence for long-range dependence in weekly stock returns of four Pacific Rim Markets (Korea, Malaysia, Singapore and New Zealand). Henry (2002) investigated the long range dependence in a sample of nine international stock index returns using parametric and nonparametric techniques. The results provide evidence of long memory in stock markets of Germany, Japan, Korea and Taiwan. Using monthly data of stock market indices of 16 OECD countries, Tolvi (2003) found statistically significant long memory for three thin stock markets (Denmark, Finland and Ireland).

3 Factor analysis

Factor models for asset returns can be divided into three types (Connor, 1995): macroeconomic, fundamental, and statistical factor models. Macroeconomic factor models (Chan, Chen and Hsieh, 1985, Chen, Roll and Ross, 1986) use economic time series indicators, such as GDP, inflation, interest rate and unemployment rate as measures of pervasive factors in asset returns. Fundamental factor models (Fama and French, 1992, 1993, 1996) use observed asset attributes, such as company size, book-to-market ratio and industrial classification to construct common factors. In statistical factor models, the common factors are extracted from the covariances of asset returns (Tsay, 2005). Connor (1995) found that statistical factor models and fundamental factor models have more explanatory power than macroeconomic factor models. An advantage of statistical factor models over fundamental factor models is the capability to identify the pervasive factors in asset returns without using any external data sources. A more recent review of factor models in capturing return comovements is given in Chan, Karceski and Lakonishok (1998).

Statistical factor analysis have been used to study the behavior of international stock returns. Drummer and Zimmermann (1992) explore the structure of 11 European stock returns using local currency stock returns. Heston et al. (1995) investigated the structure of international stock returns in Europe and the U.S., and examined the integration of capital markets using data from 6000 firms in the United States. Kraus (2001) analyzed the impact of the introduction of the euro on the return structure of European equity markets.

Standard statistical factor analysis describes the covariance relationships among observed variables in terms of a smaller number of unobserved latent variables, called factors (for details, see for instance, Everitt and Dunn, 2001 and Jonhson and Whichern, 2007). In our approach, the factors are extracted not directly from the historical returns but from their dynamic features. This procedure allows us to describe the structure of a large number of stock markets in terms of a few number of factors.

Let $y_1, y_2, ..., y_p$ be the set of the statistical characteristics of returns. The factor analysis model assumes the form

$$y_i = \theta_{i1}F_1 + \theta_{i2}F_2 + \cdots + \theta_{iq}F_q + u_i, \;\; i = 1, ..., p, \tag{9}$$

where $F_1, F_2, ..., F_q$ are unobserved latent variables or common factors, θ_{ij} is the *factor loading* of the ith variable on the jth factor, and u_i is the error or *specific factor* of the ith variable. We assume that the specific errors are uncorrelated with each other and with the common factors $F_1, F_2, ..., F_q$. The variance of the ith variable is given by

$$\sigma_i^2 = h_i^2 + \psi_i, \tag{10}$$

where $h_i^2 = \theta_{i1}^2 + \cdots + \theta_{iq}^2$ is the ith *communality* and represents the portion of the variance of the ith variable shared with the other variables via the q common factors, and ψ_i is the remaining portion of the variance of the ith variable, called the *uniqueness* or *specific variance.*

We use the classic principal-component factor analysis method in the estimation of the factor loadings and communalities, which uses the square multiple correlations as estimates of the communalities to compute the factor loadings (for a detailed discussion, see Johnson and Whichern, 2007). This procedure drops factors with eigenvalues below 1 (Kaiser criterion). We then perform an orthogonal rotation of factors through the Varimax method to simplify the factor structure. The goal of this method is to obtain factors with a few large loadings and as many loadings close to zero as possible. Factor loadings greater than 0.5 (in absolute value) are considered significant for factor interpretation purposes (Hair et al., 2006). An acceptable factor solution occurred when all variables have a significant loading on a factor and no variable has more than one significant loading. The estimated rotated factor loadings are used to compute the factor scores of each individual observation, using the regression scoring method (see Johnson and Whichern, 2007, p. 516-517). Factor scores are standardized to have zero mean and unit variance.

4 Data and exploratory analysis

The data used in this study consists of daily free float-adjusted market capitalization equity indices of developed and emerging stock markets, constructed by Morgan Stanley Capital International (MSCI). The MSCI market classification consists of following three criteria: economic development, size and liquidity, and market accessibility. The dataset includes 23 markets classified as developed[1] and 23 markets classified as emerging[2]. The data, expressed in terms of the US dollar, cover a period from January 1995 to December 2009, in a total of 3914 daily observations.

Tables 1 and 2 present features of developed and emerging market daily percentage returns under study: mean (*mean*), standard deviation (*stdev*), skewness (skew) and kurtosis (*kurt*) of log returns; Ljung-Box modified Q-statistics for the hypothesis of no autocorrelations up to order m in the returns (*qstat*) and absolute returns (*qstat2*), where m is the largest integer less or equal to $\ln(n)$; estimated long-memory d parameter of absolute returns (d), based on the log-periodogram regression method (Geweke and Porter-Hudak, 1983).

As expected, average percentage rate of return and unconditional volatility (as measured by standard deviation) for emerging markets (0.024 and 2.061) are higher than those for developed markets (0.019 and 1.548). The best performing markets were Egypt and Russia, which achieved an average percentage rate of return of 0.053. In contrast, the worst performing market was Thailand, which achieved an average percentage rate of return of -0.023. In terms of unconditional volatility, 11 of 23 emerging markets recorded daily standard deviations greater than 2% (Argentina, Brazil, China, Hungary, Indonesia, Korea, Mexico, Poland, Russia, Thailand and Turkey), while only one developed market (Finland) exceeded a standard deviation of 2%.

Almost all developed and emerging stock markets (the exceptions are Hong-Kong, Japan, Singapore, Sweden, Switzerland, China, Korea, Mexico, Philippines, and Thailand) exhibit a negative skewness, indicating that the distribution of those return indices have long left tails. The highest negative skewness coefficients correspond to stock markets (Malaysia, Indonesia, Argentina and Ireland) which exhibit as well the

[1] The developed market country indices are Australia (AUST), Austria (AUS), Belgium (BEL), Canada (CAN), Denmark (DEN), Finland (FIN), France (FRA), Germany (GER), Greece (GRE), Hong Kong (HK), Ireland (IRE), Italy (ITA), Japan (JAP), Netherlands (NET), New Zealand (NZ), Norway (NOR), Portugal (POR), Singapore (SING), Spain (SPA), Sweden (SWE), Switzerland (SWI), United Kingdom (UK) and United States (US).

[2] The emerging market country indices are Argentina (ARG), Brazil (BRA), Chile (CHI), China (CHI), Czech Republic (CR), Colombia (COL), Egypt (EGY), Hungary (HUN), India (IND), Indonesia (INDO), Israel (ISR), Korea (KOR), Malaysia (MAL), Mexico (MEX), Morocco (MOR), Peru (PER), Philippines (PHI), Poland (POL), Russia (RUS), South Africa (SA), Taiwan (TAI), Thailand (THA) and Turkey (TUR).

highest excess of kurtosis (68.22, 28.22, 20.63 and 16.68, respectively). The lowest kurtosis coefficient correspond to stock market of Taiwan (5.58). In general, the emerging market returns exhibit more excess kurtosis than developed market returns.

According to the Ljung-Box test statistic for serial correlation in the returns (*qstat*), all but four (Hong-Kong, New Zealand, Argentina and Israel) countries show significant evidence at the 1% level of short-term linear dependence in the return series. On the other hand, the Ljung-Box test statistic for serial correlation in the absolute returns (*qstat2*) indicate the presence of nonlinear dependence and apparent conditional heteroskedasticity effects for all return series. In general, emerging market returns seem to have stronger linear dependence than developed market returns. By contrast, the nonlinear dependence behavior is more salient in developed market returns.This can be explained by the fact that the volatility in emerging markets is primarily driven by local factors (Bekaert and Harvey, 1997).

The results of the Geweke and Porter-Hudak estimates of d suggest that the hypothesis of no long memory is rejected at the 5% level in all but two (Argentina and Colombia) market returns under study. On the other hand, in 12 of 23 developed markets and 7 of 23 emerging markets there is strong evidence of long memory, with d estimates suggesting that absolute returns are in a nonstationary region ($d > 0.5$). In particular, very strong evidence of long memory can be found in the absolute returns of the markets of Austria, Norway, Indonesia and Korea. Average long memory d estimate of absolute returns in developed markets (0.50) is similar to that in emerging markets (0.47).

5 Statistical results

5.1 Factor loadings

To investigate how the structure of the global stock market returns has evolved in the period under study, we divided the entire sample period into three sub-sample periods covering 1995:01-1999:12, 2000:01-2004:12 and 2005:01-2009:12. We apply principal-component factor analysis separately for each of the 5-year periods, and we obtain a factor solution for correlations of the 7 statistical variables (*mean*, *stdev*, *skew*, *kurt*, *qstat*, *qstat2* and d).

In order to identify clusters of markets and possible multivariate outliers, we compute scores for the first two factors derived from factor analysis. We identify factor scores have values greater than ± 2 as outliers. In our analysis, we classify as outliers the markets of Colombia, Malaysia, Indonesia, Korea and Thailand in the period of 1995-1999,

Table 1: Statistical features for developed stock market returns

Market	*mean*	*stdev*	*skew*	*kurt*	*qstat*	*qstat*2	*d*
Australia	0.027	1.465	-0.848	14.52	20.6*	3056.6*	0.581*
Austria	0.010	1.576	-0.297	14.42	27.8*	4265.5*	0.638*
Belgium	0.010	1.427	-0.553	12.88	77.7*	2851.1*	0.501*
Canada	0.039	1.444	-0.841	13.40	66.0*	2743.4*	0.568*
Denmark	0.038	1.379	-0.377	11.27	45.6*	2283.8*	0.416*
Finland	0.033	2.315	-0.355	9.12	23.7*	1329.9*	0.614*
France	0.025	1.479	-0.042	10.13	67.7*	1955.0*	0.393*
Germany	0.021	1.578	-0.071	8.27	28.9*	2077.1*	0.553*
Greece	0.017	1.853	-0.128	7.58	55.2*	1512.1*	0.401*
Hong-Kong	0.015	1.673	0.036	12.03	18.1**	1801.7*	0.504*
Ireland	-0.008	1.659	-0.890	16.68	48.6*	3019.9*	0.589*
Italy	0.016	1.514	-0.015	9.65	80.1*	1898.0*	0.442*
Japan	-0.011	1.509	0.125	7.08	24.7*	890.0*	0.362*
Netherlands	0.019	1.469	-0.157	10.00	78.2*	2676.4*	0.402*
Norway	0.026	1.811	-0.519	12.47	38.4*	3577.5*	0.704*
New Zealand	0.000	1.455	-0.532	11.13	18.5**	1287.2*	0.573*
Portugal	0.019	1.247	-0.167	12.10	76.9*	1327.3*	0.417*
Singapore	0.008	1.527	0.024	9.33	28.4*	1907.9*	0.464*
Spain	0.042	1.524	-0.082	9.75	50.3*	2022.2*	0.436*
Sweden	0.034	1.874	0.062	7.93	40.3*	2029.2*	0.545*
Switzerland	0.030	1.234	0.020	8.41	63.4*	1695.1*	0.378*
United Kingdom	0.015	1.326	-0.127	13.35	105.2*	2728.0*	0.442*
United States	0.023	1.259	-0.216	11.52	41.7*	2511.9*	0.549*
Average	0.019	1.548	-0.259	11.00	48.95	2236.8	0.499

Notes: mean (*mean*), standard deviation (*stdev*), skweness (*skew*) and kurtosis (*kurt*) of log returns; Ljung-Box modified Q-statistics for the hypothesis of no autocorrelations up to order m in the returns (*qstat*) and absolute returns (*qstat2*); long-memory d parameter of absolute returns (d), where the number of periodogram ordinates (l) used in the Geweke and Porter-Hudak regression (Geweke and Porter-Hudak, 1983) is given by $l = n^{0.5}$.
* (**) indicates rejection of the null hypothesis at the 1% (5%) level.

Table 2: Statistical features for emerging stock market returns

Market	*mean*	*stdev*	*skew*	*kurt*	*qstat*	*qstat2*	*d*
Argentina	0.014	2.450	-1.092	20.63	14.0	962.2*	0.322*
Brazil	0.041	2.491	-0.091	10.02	48.0*	2077.9*	0.470*
Chile	0.019	1.358	-0.064	14.66	115.3*	1648.1*	0.360*
China	-0.002	2.119	0.040	8.23	67.3*	1649.1*	0.480*
Colombia	0.042	1.668	-0.127	13.32	175.4*	1958.6*	0.353*
Czech Republic	0.043	1.783	-0.166	14.34	63.2*	2064.4*	0.349*
Egypt	0.053	1.653	-0.338	10.23	66.0*	702.4*	0.462*
Hungary	0.051	2.239	-0.301	13.05	74.6*	1651.7*	0.379*
India	0.030	1.821	-0.055	9.78	54.2*	1272.8*	0.460*
Indonesia	0.008	2.943	-1.046	28.22	113.7*	2453.4*	0.642*
Israel	0.033	1.462	-0.355	7.89	16.8**	848.2*	0.512*
Korea	0.015	2.587	0.221	15.59	110.3*	1844.1*	0.681*
Malaysia	0.001	1.904	-0.852	68.22	100.6*	2366.5*	0.466*
Mexico	0.038	2.038	0.019	14.45	35.4*	1506.4*	0.459*
Morocco	0.036	0.963	-0.127	8.43	166.8*	1170.0*	0.516*
Peru	0.049	1.836	-0.152	10.22	48.2*	1806.2*	0.548*
Philippines	-0.021	1.809	0.534	15.30	135.1*	994.5*	0.456*
Poland	0.021	2.110	-0.158	6.63	52.0*	1086.2*	0.444*
Russia	0.053	3.259	-0.366	12.62	37.0*	1879.4*	0.475*
South Africa	0.020	1.765	-0.476	9.08	44.9*	1886.0*	0.536*
Taiwan	-0.005	1.757	-0.069	5.58	31.9*	524.7*	0.491*
Thailand	-0.023	2.229	0.495	12.60	121.7*	1238.2*	0.524*
Turkey	0.042	3.164	-0.145	9.14	29.5*	847.7*	0.320*
Average	0.024	2.061	-0.203	14.71	74.86	1497.3	0.465

Notes: As in Table 1.

the markets of Colombia, Philippines, Morocco and Argentina in the period of 2000-2004, and the market of Morocco in the period 2005-2009. Outliers can impact correlations strongly and change factor structure in the solution. Thus, we investigate whether communalities and factor loadings change in the factor solution by omitting the countries that are considered outliers. The factor analysis results suggest that outliers have some impact on factor structure, especially over the first and second 5-year periods (1995-1999 and 2000-2005). Therefore, the factor analysis solution without outliers is used for interpretation purposes. The factor loadings are then transformed through the Varimax rotation. The two sets of unrotated and rotated loadings are given in Table 3.

For the period 1995-1999, the factor analysis method retained 3 factors with an eigenvalue of 1 or greater. The cumulative variance accounted by these three factors is 4.811, which is about 69.7% (4.881/7) of the total variance. The factor 1 in the unrotated solution accounts for 32.4% (2.264/7) of the total variance and 46.4% (2.264/4.881) of the common variance, the factor 2 accounts for 22.8% of the total variance and 37.8% of the common variance, and the factor 3 accounts for 14.5% of the total variance and 20.9% of the common variance. The communalities indicate the amount of variance that each variable shares with all other variables in the set. All variables have communality estimates greater than 0.5, and 4 of the 7 variables (*kurt*, *qstat*, *qstat2* and d) have communality estimates greater than 0.7, which means that these variables are highly correlated with the retained factors.

Using the threshold of ± 0.5 for identifying significant loadings, we can see that all variables in the unrotated solution have a significant loading on a factor. However, *kurt* has significant cross-loadings on the first two retained factors and stdev does not load significantly on any factor. The Varimax rotation improved the factor structure. After the rotation, *kurt* loads uniquely on factor 3 and *stdev* loads on factor 2. However, *qstat2* now loads significantly both on factor 1 and factor 2. Nevertheless, the rotated solution is used to interpret the factors. Factor 1 has three variables with significant loadings (*mean*, *qstat* and *qstat2*), factor 2 has three variables with significant loadings (*stdev*, *kurt* and *qstat2*), and factor 3 has two (*skew* and d). The pattern of factor loadings on factor 1 indicates that short-term linear dependence and short-term nonlinear dependence are positively related and these measures are negatively related with mean return. This factor seems to represent the volatility clustering of the return series. In factor 2, unconditional volatility, excess kurtosis and nonlinear dependence load positively on the factor. This factor is characterized by the distributional properties of returns. In factor3, skewness and long-range dependence have oppo-

Table 3: Factor analysis for empirical properties of global stock market returns

Period I: 1995-1999

	Unrotated factors			Rotated factors			
Variable	F1	F2	F3	F1	F2	F3	Communality
mean	-0.679	0.181	-0.314	-0.746	-0.180	0.055	0.59
stdev	0.497	0.497	-0.243	0.048	0.742	-0.032	0.55
skew	0.328	-0.664	-0.228	0.365	-0.201	-0.653	0.60
kurt	0.511	0.576	-0.361	-0.037	0.845	-0.079	0.72
qstat	0.648	-0.480	0.346	0.853	-0.065	-0.196	0.77
qstat2	0.812	0.324	0.220	0.604	0.656	0.133	0.81
d	-0.330	0.461	0.714	-0.020	-0.107	0.905	0.83
Eigenvalue	2.264	1.599	1.018	1.786	1.783	1.312	
Proportion	0.324	0.228	0.145	0.255	0.255	0.188	

Period II: 2000-2004

	Unrotated factors			Rotated factors			
Variable	F1	F2	F3	F1	F2	F3	Communality
mean	-0.655	-0.190	-0.213	-0.585	-0.405	-0.064	0.51
stdev	0.368	0.685	0.023	0.178	0.712	-0.257	0.61
skew	0.533	-0.513	-0.029	0.648	-0.354	0.099	0.55
kurt	0.106	0.781	0.352	-0.097	0.857	0.041	0.75
qstat	-0.114	-0.386	0.885	-0.002	-0.050	0.971	0.94
qstat2	0.760	-0.181	0.211	0.783	0.095	0.181	0.65
d	0.793	-0.117	-0.289	0.793	-0.023	-0.310	0.73
Eigenvalue	2.078	1.574	1.081	2.044	1.537	1.153	
Proportion	0.297	0.225	0.154	0.292	0.220	0.165	

Period III: 2005-2009

	Unrotated factors			Rotated factors			
Variable	F1	F2	F3	F1	F2	F3	Communality
mean	-0.566	0.452	0.273	-0.432	0.635	0.098	0.60
stdev	-0.092	0.849	0.127	0.154	0.849	-0.024	0.75
skew	0.302	0.319	-0.663	0.461	0.085	-0.643	0.62
kurt	0.424	0.021	0.757	0.319	0.054	0.806	0.75
qstat	0.539	-0.230	0.117	0.427	-0.341	0.242	0.36
qstat2	0.820	0.071	0.002	0.797	-0.158	0.131	0.68
d	0.786	0.374	-0.024	0.859	0.131	0.053	0.76
Eigenvalue	2.181	1.224	1.118	2.079	1.292	1.152	
Proportion	0.312	0.175	0.160	0.297	0.185	0.165	

site signs. This factor might represent the long-memory behavior of the return series.

Examining the factor analysis solution, using data for the period 2000-2004, we found that the method retained 3 factors with eigenvalues greater than 1. The three factors retained represent 67.6% of the total variance of the 7 variables. All of the communalities are above 0.50 meaning that all the variables share more than one-half of their variance with the three retained factors. In particular, *kurt*, *qstat* and *d* variables have a communality greater than 0.7. In the unrotated factor solution, the percentages of the total variance explained by each of the three factors are 29.7%, 22.5% and 15.4%, respectively. One of the 7 variables (*skew*) has significant loadings on two factors (factor 1 and factor 2) but with opposite signs. The factor 1 has four significant loadings for variables *mean*, *skew*, *qstat2* and *d*, the factor 2 has significant loadings for variables *stdev*, *skew* and *kurt*, and the factor 3 has one significant loading for variable *qstat*. In the Varimax rotated factor solution, the three factors account for 29.2%, 22.0% and 16.5% of the total variance, respectively. Now, the variable *skew* for each common factor has no cross-loadings on the other common factors. Factor 1 has three variables with significant positive loadings (*skew*, *qstat2* and *d*) and one variable with significant negative loading (*mean*). Factor 2 has two significant positive loadings (*stdev* and *kurt*), and factor 3 has one significant loading (*qstat*). The rotated factor solution seems to be more meaningful in terms of the theoretical interpretation of its factors. Factor 1 is dominated by nonlinear (short and long-range) dependence. In general, the presence of nonlinear dependence and apparent conditional heteroskedasticity effects is more salient in markets with higher long-memory behavior. Factor 2 seems to be associated to the shape of the return distribution and to deviations from the standard normal distribution. Factor 3 seems to represent the serial dependence of price changes.

For the most recent period (2005-2009), the proportion of the total variance explained by the three-factor solutions in about 64.6%. The Varimax rotation did not improve the factor structure. In fact, after the rotation qstat has no significant loadings on any of the factors. Factor 1 seems to represent the nonlinear dependence of stock market returns. The short-range and long-range dependence measures (*qstat2* and *d*, respectively) load highly on this factor. The standard deviation of returns has a large positive loading on factor 2. The variables skew and kurt have high loadings on factor 3, but with opposite signs.

5.2 Factor scores

Figure 1 shows bi-dimensional plots of stock market scores given by the two factors that explain the largest proportion of the total variance. The plot in the top of Figure 1 corresponds to the period 1995-1999. It can be seen that most developed markets form a cluster with negative scores on factor 1 and rather low scores on factor 2. The exceptions are the markets of Greece, Japan and Norway, with scores on factor 1 close to zero, and the Pacific Rim markets of Hong-Kong, New Zealand and Singapore, with positive scores on factor 1. In this period, the markets of Hong Kong and Singapore experienced rather high values of the short-term dependence parameters (*qstat* and *qstat2*), which load positively on factor 1. On the other hand, the market of New Zealand exhibited the highest coefficient of kurtosis (also loading positively on factor 1) amongst the developed markets group. While developed markets are closely clustered, emerging markets are widely scattered both in terms of factor 1 and factor 2, showing a richer diversity of dynamic behaviors, as measured by the empirical properties of the return series. Nevertheless, emerging markets generally display positive values on factor 1, which are predominantly determined by high levels of short-term dependence. In terms of factor 2, the scores for emerging markets range from large negative values (e.g., the North African markets of Egypt and Morocco) to large positive values (e.g., the East European markets of Hungary and Russia).

The plot in the middle of Figure 1 shows the results for the period 2000-2004. This period is rather atypical in the sense that there is no clear separation between developed and emerging markets. It can be seen that many Western markets exhibit positive scores on factor 1. These scores are mostly driven by non-linear short-term and long-term dependencies. In particular, the market of the Netherlands experienced the largest values of *qstat2* and d amongst the 46 markets. Nonetheless, many developed markets, such as Australia, New Zealand, Norway, Australia and Ireland, exhibit negative scores on factor 1. For instance, the market of Norway had the lowest value of the long memory parameter d in the sample. With respect to factor 2, the markets of Finland, Turkey and Indonesia clearly stand apart from the remaining markets. The large positive scores given by factor 2 are primarily determined by large values of unconditional volatility and kurtosis. In fact, the market of Finland had the largest values of *stdev* and *kurt* amongst the group of developed markets, while the markets of Turkey and Indonesia exhibited the largest values of these parameters in the group of emerging markets.

Finally, the plot in the bottom of Figure 1 shows the factor scores for the period 2005-2009. As in the first period, most European markets are tightly clustered and close to the North American markets of Canada and the Unites States. Also, the Pacific Rim markets of Australia, New Zealand and Japan can be found in this cluster. All developed markets score positively on factor 1, with the exception of Greece. On the other hand, many emerging markets have negative scores on factor 1, with the exceptions of the markets of the Czech Republic, Hungary, Korea, Mexico and Russia. Again, emerging markets are more scattered than their developed counterparts. In particular, markets which experienced higher levels of volatility during this period, such as Brazil, Hungary, Russia and Turkey, are located far from those that had lower volatilities, such as Israel and Malaysia.

6 Conclusions

In this article, we employ a factor analysis approach to examine the structure of returns across 46 international stock markets over the period 1995-2009. Common factors are extracted not from the historical returns but from their empirical properties by principal-component factor analysis. This procedure allows us to describe the structure of a large number of stock markets in terms of a few number of factors, without imposing any restriction about the time dependence of the observations. The estimated factor loadings were then used to generate scoring coefficients of each of these factors for each country. In order to grasp the stableness of the empirical findings across time, the analysis was performed on three datasets covering periods of five years.

The factor analysis reveals that the dependence structure of stock market returns differs substantially between developed and emerging markets. However, this structure has not been constant over the time period covered by the data. For the period 1995-1999, the principal factor is strongly and positively correlated with short-term (linear and nonlinear) dependence in returns and, on the other hand, it is highly and negatively correlated with the mean return. The factor scores derived from these patterns are found to be positive for most emerging markets and negative for most developed markets. In fact, developed and emerging markets form two reasonably well separated clusters. Somewhat different results are obtained in the period 2000-2004. The principal factor is highly and positively correlated with the nonlinear dependence properties but negatively correlated with the mean return. When compared to the two adjacent time periods, this period is rather atypical since there are no clearly separated clusters for developed and emerging markets. During the most recent period (2005-2009), the principal factor is positively correlated with short- and long-term dependence parameters and negatively correlated with mean return. In contrast with the

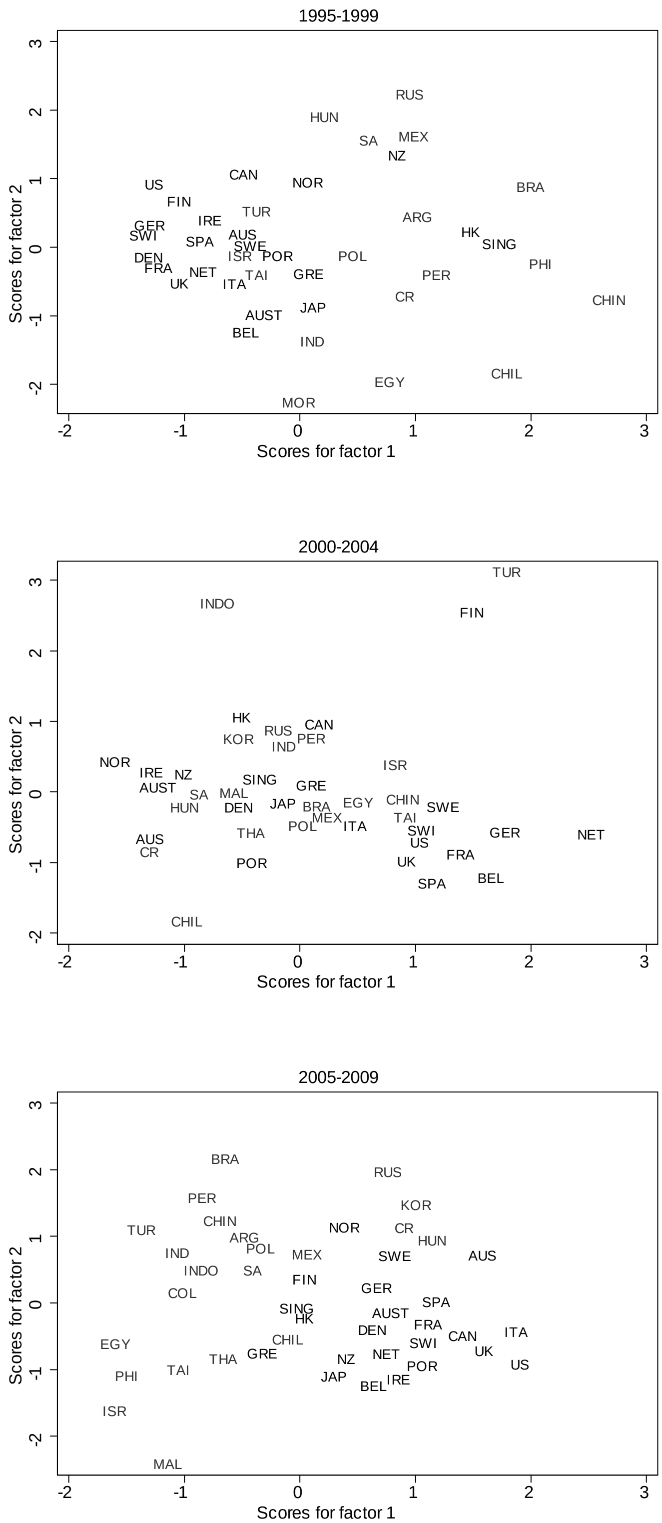

Figure 1: Score plots of the two principal factors for stock market returns (emerging markets denoted by gray color)

first period, the factor scores are positive for most developed markets and negative for most emerging markets. In consonance with the first period, two reasonably well separated clusters can be found.

Irrespectively of the analyzed period, the factor analysis invariably produces a factor that loads significantly on the mean return and at least on two of the three parameters that describe short- and long-term serial dependences. This factor always indicates a negative relation between mean returns and the statistics that describe the correlation structure. Furthermore, the factor analysis always produces a second factor that loads significantly on the skewness and kurtosis coefficients of the return distributions, additionally suggesting a negative relationship between these statistics. This observation corroborates the findings of Bekaert and Harvey (2002) using average monthly returns for emerging markets. Overall, the results further suggest that the empirical properties of returns across emerging markets are relatively less correlated than those across developed markets. This can be perceived by the higher dispersion of emerging markets in the score plots and by the outliers identified in the analysis, which always belong to the emerging markets group.

Acknowledgment: The authors would like to thank Ana Diniz, Luís Martins and Rui Menezes for their helpful suggestions and contributions. This research was supported by a grant from the Fundação para a Ciência e a Tecnologia.

References

[1] Baillie, R. T. (1996). "Long memory processes and fractional integration in econometrics", *Journal of Econometrics*, 73, 5-59.

[2] Bastos, J. A. and Caiado, J. (2009). "Clustering global equity markets with variance ratio tests", CEMAPRE Working Paper n.^{o} 0904.

[3] Bekaert, G. and Harvey, C. (1997). "Emerging equity market volatility", *Journal of Financial Economics*, 43, 29-77.

[4] Bekaert, G. and Harvey, C. (2002). "Research in Emerging Markets Finance: Looking to the Future", *Emerging Markets Review*, 3, 429-448

[5] Bekaert, G., Hodrick, R. J. and Zhang, X. (2010). "International stock return comovements", *Journal of Finance*, 64, 2591-2626.

[6] Beran, J. A. (1994). *Statistics for Long-Memory Processes*, Chapman & Hall, New York.

[7] Bonanno, G., Caldarelli, G., Lillo, F., Miccieché, S., Vandewalle N. and Mantegna, R. (2004). "Networks of equities in financial markets", *European Physical Journal B*, 38, 363-371.

[8] Caiado, J. and Crato, N. (2007). "A GARCH-based method for clustering of financial time series: International stock markets ev-

idence", in: *Recent Advances in Stochastic Modeling and Data Analysis* (Edited by C. H. Skiadas), World Scientific Publishing, 542-551.

[9] Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1996). *The Econometrics of Financial Markets*, Princeton University Press, New Jersey

[10] Chow, K. V., Pan, M. and Sakano, R. (1996). "On the long-term or short-term dependence in stock prices: Evidence from international stock markets", *Review of Quantitative Finance and Accounting*, 6, 181-194.

[11] Chan, K. C., Chen, N. F. and Hsieh, D. A. (1985). "An exploratory investigation of the firm size effect", *Journal of Financial Economics*, 14, 451-471.

[12] Chan, L. K. C., Karceski, J. and Lakonishok, J. (1998). "The risk and return from factors", *Journal of Financial and Quantitative Analysis*, 33, 159-188.

[13] Chen, N. F., Roll, R. and Ross, S. A. (1986). "Economic forces and the stock market", *Journal of Business*, 59, 383-403.

[14] Connor, G. (1995). "The three types of factor models: A comparison of their explanatory power", *Financial Analysts Journal*, 51, 42-46.

[15] Connor, G. and Korajczyk, R. (1988). "Risk and return in an equilibrium APT: Application of a new test methodology", *Journal of Financial Economics*, 15, 373-394.

[16] Cont, R. (2001). "Empirical properties of asset returns: stylized facts and statistical issues", *Quantitative Finance*, 1, 223-236.

[17] Crato, N. (1994). "Some international evidence regarding the stochastic behavior of stock returns", *Applied Financial Economics*, 4, 33-39.

[18] Ding, Z., Granger, C. W. J. and Engle, R. F. (1993). "A long memory property of stock market returns and a new model", *Journal of Empirical Finance*, 1, 83-106.

[19] Drummen, M. and Zimmerman, H. (1992). "The structure of European stock returns", *Financial Analysts Journal*, 48, 15-26.

[20] Engle, R. (1982). "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation", *Econometrica*, 50, 987-1008.

[21] Engle, R. and Kroner. K. (1995). "Multivariate simultaneous generalized ARCH", *Econometric Theory*, 11, 122-150.

[22] Engle, R. and Susmel, R. (1993). "Common volatility in international equity markets", *Journal of Business & Economic Statistics*, 11, 167-176.

[23] Everitt, B. and Dunn, G. (2001). *Applied Multivariate Data Analy-*

sis, 2th Ed., Hodder Arnold, London.

[24] Fama, E. F. and French, K. R. (1992). "The cross-section of expected stock returns", *Journal of Finance*, 47, 427-465.

[25] Fama, E. F. and French, K. R. (1993). "Common risk factors in the returns on stocks and bonds", *Journal of Financial Economics*, 33, 3-56.

[26] Fama, E. F. and French, K. R. (1996). "Multifactor explanations of asset pricing anomalies", *Journal of Finance*, 51, 55-84.

[27] Franses, P. H. and van Dijk, D. (2000). *Non-Linear Time Series Models in Empirical Finance*, Cambridge University Press, Cambridge.

[28] Geweke, J. and Porter-Hudak, S. (1983). "The estimation and application of long-memory time series models", *Journal of Time Series Analysis*, 1, 15-29.

[29] Granger, C. W. J. and Ding, Z. (1996). "Varieties of long-memory models", *Journal of Econometrics*, 73, 61-77.

[30] Hair, J. F., Black, W. C., Babin, B. J. and Anderson, R. E. (2006). *Multivariate Data Analysis*, 6th Ed., Prentice-Hall, New Jersey.

[31] Harvey, C. (1995). "Predictable risk and returns in emerging markets", *The Review of Financial Studies*, 8, 773-816.

[32] Henry, O. T. (2002). "Long memory in stock returns: Some international evidence", *Applied Financial Economics*, 12, 725-729.

[33] Heston, S. L., Rouwenhorst, K. G. and Wessels, R. E. (1995). "The structure of international stock returns and the integration of capital markets", *Journal of Empirical Finance*, 2, 173-197.

[34] Hui, T. (2005). "Portfolio diversification: a factor analysis approach", *Applied Financial Economics*, 15, 821-834.

[35] Johnson, R. and Wichern, D. (2007). *Applied Multivariate Statistical Analysis*. 6th Ed., Prentice-Hall, New Jersey.

[36] Karolyi, G. and Stulz, R. (1996). "Why do markets move together? An investigation of U.S.-Japan return comovements", *The Journal of Finance*, 51, 951-986.

[37] King, M., Sentana, E. and Wadhwani, S. (1994). "Volatility and links between national stock markets", *Econometrica*, 62, 901-933.

[38] Kraus, T. (2001). "The impact of the EMU on the structure of European equity returns: An empirical analysis of the first 21 months", IMF Working Paper 01/84, International Monetary Fund.

[39] Limam, I. (2003). "Is long memory a property of thin stock markets? International evidence using Arab countries", *Review of Middle East Economics and Finance*, 1, 251-266.

[40] Lin, W., Engle, R. and Ito, T. (1994). "Do bulls and bears move across borders? International transmission of stock returns and

volatility", *The Review of Financial Studies*, 7, 507-538.

[41] Lo, A. W. (1991). "Long-term memory in stock market prices", *Econometrica*, 59, 1279-1313.

[42] Lo, A.W. and Mackinlay, A. C. (1988). "Stock market prices do not follow random walks: evidence from a simple specification test", *The Review of Financial Studies*, 1, 41-66.

[43] Lobato, I. N. and Savin, N.E. (1998). "Real and Spurious Long-Memory Properties of Stock-Market Data", *Journal of Business & Economic Statistics*, 16, 261-268.

[44] Longin, F. and Solnik, B. (1995). "Is the correlation in international equity returns constant: 1960-1990?", *Journal of International Money and Finance*, 14, 3-26.

[45] Morana, C. and Beltratti, A. (2008). "Comovements in international stock markets", *Journal of International Financial Markets, Institutions and Money*, 18, 31-45.

[46] Panton, D. B., Lessig, V. P. and Joy, O. M. (1979). "Comovement of international equity markets: A taxonomic approach", *The Journal of Financial and Quantitative Analysis*, 11, 415-432.

[47] Poterba, J. and Summers, L. (1988). "Mean reversion in stock returns: Evidence and implications", *Journal of Financial Economics*, 22, 27–60.

[48] Rita, D. and Costantini, M. (2006). "Comovements and correlations in international stock markets", *The European Journal of Finance*, 12, 567-582.

[49] Sadique, S. and Silvapulle, P. (2001). "Long-term memory in stock market returns: international evidence", *International Journal of Finance and Economics*, 6, 59-67.

[50] Syriopoulos, T. (2004). "International portfolio diversification to Central European stock markets", *Applied Financial Economics*, 14, 1253-1268.

[51] Tahai, A., Rutledge, R. and Karim, K. (2004). "An examination of financial integration for the group of seven (G7) industrialized countries using an I(2) cointegration model", *Applied Financial Economics*, 14, 327-335.

[52] Tolvi, J. (2003). "Long memory and outliers in stock market returns". *Applied Financial Economics*, 13, 495-502.

[53] Tsay, R. (2005), *Analysis of Financial Time Series*, 2nd ed., Wiley, New Jersey.

[54] Voronkova, S. (2004). "Equity market integration in Central European emerging markets: A cointegration analysis with shifting regimes", *International Review of Financial Analysis*, 13, 633-647.