Efficient market hypothesis in European stock markets

Maria Rosa Borges

Routledge
Taylor & Francis Group

Efficient market hypothesis in European stock markets

Maria Rosa Borges*

Instituto Superior de Economia e Gestão, Technical University of Lisbon and UECE (Research Unit on Complexity and Economics), Rua Miguel Lupi, 20, 1249-078 Lisboa, Portugal

This paper reports the results of tests on the weak-form market efficiency applied to stock market indexes of UK, France, Germany, Spain, Greece and Portugal, from January 1993 to December 2007. We use a runs test, and joint variance ratio tests, which are performed using daily and weekly data for the period 1993–2007 and for a subset, 2003–2007. Daily and weekly returns are not normally distributed, because they are negatively skewed and leptokurtic, and also display conditional heteroscedasticity. Overall, we find mixed evidence on the efficient market hypothesis (EMH). The hypothesis is rejected on daily data for Portugal and Greece, due to first-order positive autocorrelation in the returns. However, the empirical tests show that these two countries have been approaching a martingale behavior after 2003. France and UK data rejects EMH, due to the presence of mean reversion in weekly data, and stronger in recent years. Taken together, the tests for Germany and Spain do not allow the rejection of EMH, this last market being the most efficient.

Keywords: market efficiency; martingale; European stock markets; variance ratio test

JEL Classification: G14; G15

1. Introduction

Efficient market theory and the random walk hypothesis have been major issues in the financial literature, for the past 30 years. While a random walk does not imply that a market cannot be exploited by insider traders, it does imply that excess returns are not attainable through the use of information contained in the past movement of prices. The validity of the efficient market hypothesis (EMH) has important implications for financial theories and investment strategies, and so this issue is relevant for academicians, investors and regulatory authorities. Academicians seek to understand the behavior of stock prices, and standard risk-return models, such as the capital asset pricing model, depend on the hypotheses of normality or random walk behavior of returns. For investors, trading strategies have to be designed taking into account if future returns can be predicted based on their past behavior, or not, which would be the case if weak-form EMH is confirmed. Finally, if a stock market is not efficient, the pricing mechanism does not ensure the efficient allocation of capital within an economy, with negative effects for the overall economy. Evidence of inefficiency may lead regulatory authorities to take the necessary steps and reforms to correct it.

Since the seminal work of Fama (1970), several studies have shown that stock price returns do not follow random walks and are not normally distributed, including Fama and French (1988) and Lo and Mackinlay (1988), among many others. The globalization of markets spawned interest

on the study of this issue, with many studies both on individual markets and regional markets, such as Latin America (Urrutia 1995; Grieb and Reyes 1999; Charles and Darné 2009), Africa (Magnusson and Wydick 2002; Smith, Jefferis, and Ryoo 2002), Asia (Huang 1995; Groenewold and Ariff 1998; Kim and Shamsuddin 2008a), Middle East (Abraham, Seyyed, and Alsakran 2002; Al-Khazali, Ding, and Pyun 2007), Europe (Smith and Ryoo 2003; Worthington and Higgs 2004; Smith 2009), and even comprehensive world-wide studies (Kim and Shamsuddin 2008b), several reporting non-conformity with weak-form EMH, more evidently in emerging markets, as expected. The list is too extensive for a comprehensive survey, which is beyond the purpose of this study.

In this paper, we study the weak-from EMH in six European stock markets, defined as developed both by the FTSE country classification criteria and the Standard & Poor's/IFC criteria. The choice of developed stock markets stems from the fact that these markets should be precisely the ones where it would be expected EMH to hold true. A rejection of EMH, in highly developed markets, casts strong doubts over the theory behind the hypothesis, while a rejection of EMH in emerging and illiquid markets is generally simply taken as evidence that those markets are still inefficient, and not as evidence that the theory behind the hypothesis may be flawed. EMH should be tested against the *a priori* most developed and liquid markets. This is the reason why the four largest European stock markets, UK, France, Germany and Spain, are included in this study. For comparison, two of the European countries whose markets have smaller capitalization and where development has occurred more recently, Greece and Portugal, are also included. Note that Portugal and Greece were classified as emerging markets by Standard & Poor's/IFC until April 1999 and May 2001, respectively.

This paper contributes to the literature on EMH in several aspects. First, the data covers very recent years, up to 2007, which have not been covered in previous studies of Western European developed markets. Smith and Ryoo (2003) use data until 1998, Worthington and Higgs (2004) until 2003 and Smith (2009) until 2007, but this last study only covers Eastern European emerging markets. Second, the results are obtained both for 1993–2007 and for 2003–2007. The comparison of two periods is useful (even if they are overlapping) in assessing if higher efficiency is the consequence of a development process, under which (most) markets are gradually becoming more efficient. Third, this study uses some of the most recent statistical techniques, which are more powerful in detecting departures from EMH, and have not been applied in the vast majority of previous studies. These include wild bootstrapping of joint variance ratio (VR) tests (as in Kim 2006) and joint signs-based VR tests (as in Kim and Shamsuddin 2008a, 2008b). Finally, as in Worthington and Higgs (2004), this study includes a runs test, thus diversifying the types of tests and reducing the risk that a spurious result from one of the tests might affect the conclusions.

The remainder of the paper is organized as follows. The next section explains the methodology of the different statistical tests used to detect departures from EMH. Section 3 presents the data. Section 4 presents the results from the statistical tests. Section 5 compares the results with previous studies on the same European markets. Conclusions are drawn in Section 6.

2.   Methodology

Weak form market efficiency implies that prices of securities traded in the market cannot be predicted by using historical price information. In turn, this implies that prices in such a market are serially uncorrelated. The random walk hypothesis posits that successive price changes are

random, and is in practice very restrictive, because it implies that in the process

$$X_t = \phi X_{t-1} + \varepsilon_t \tag{1}$$

the error term is an independent and identically distributed (i.i.d.) sequence. However, it is possible to relax the assumption of i.i.d. returns within weak form efficiency. X_t is a martingale if

$$E[X_{t+1}|\{X_t, X_{t-1}, \dots .\}] = X_t, \tag{2}$$

where ε_t is a martingale difference sequence. This condition means that the present and past values of X_t are useless to improve forecasts about X_{t+1}. If X_t is the log of a financial index, the returns are not predictable, but they are not necessarily i.i.d., and can display, for example, conditional heteroscedasticity.

Some of the commonly used tests of EMH assume i.i.d. returns. However, we show in Section 3 that all the daily return series and most of the weekly return series in our data display autoregressive conditional heteroscedasticity, and so i.i.d.-dependent tests are not adequate. Therefore, we only use tests that remain valid under the presence of ARCH effects in the data, thus testing the martingale hypothesis.

2.1 Runs test

To test for serial independence in the returns, we employ a runs test, which determines whether successive price changes are independent of each other, as should happen under EMH. By observing the number of runs, that is, the successive price changes (or returns) with the same sign, in a sequence of successive price changes (or returns), we can test that null hypothesis. We classify each return according to its position with respect to the mean return of the period under analysis. We have a positive sign $(+)$ each time the return is above the mean return and a negative sign $(-)$ if it is below the mean return, thus allowing for an eventual time drift in the series of returns. Note that this is a non-parametric test, which does not require the returns to be normally distributed, and so is a martingale test. The runs test is based on the premise that if price changes (returns) are random, the actual number of runs (R) should be close to the expected number of runs (μ_R).

Let n_+ and n_- be the number of positive returns $(+)$ and negative returns $(-)$ in a sample with n observations, where $n = n_+ + n_-$. For large sample sizes, the test statistic is approximately normally distributed:

$$Z = \frac{R - \mu_R}{\sigma_R} \approx N(0, 1), \tag{3}$$

where $\mu_R = 2n_+n_-/n + 1$ and $\sigma_R = \sqrt{2n_+n_-(2n_+n_- - n)/n^2(n - 1)}$.

2.2 VR tests

2.2.1 VR tests based on return values
An important property of the random walk is explored by the final tests, the VR tests. Let y_t be an asset return at time t, where $t = 1, \dots, T$. The ratio of the variance of the kth difference scaled by k to the variance of the first difference tends to equal one, that is,

$$\mathrm{VR}(k) = \frac{\sigma^2(k)}{\sigma^2(1)}, \tag{4}$$

where $\sigma^2(k)$ is $1/k$ the variance of the kth difference and $\sigma^2(1)$ is the variance of the first difference. Under the null hypothesis of a random walk, $\text{VR}(k)$ must approach unity. If this ratio is less than 1 at long horizons, we have indication of negative serial correlation (mean-reversion) and ratios greater than 1 at long horizons indicate positive serial correlation (mean-aversion or persistence).

Lo and Mackinlay (1988) propose two test statistics that explore this property. Define the estimator of the variance of the k-period difference, $\sigma^2(k)$, as

$$\sigma^2(k) = \frac{1}{Tk} \sum_{t=k}^{Tq} (y_t + \cdots + y_{t-k+1} - k\hat{\mu})^2 \tag{5}$$

where $\hat{\mu} = (1/t) \sum_{t=1}^{T} y_t$, and the estimator of the variance of the first difference, $\sigma^2(1)$, as

$$\sigma^2(1) = \frac{1}{T} \sum_{t=k}^{Tq} (y_t - \hat{\mu})^2. \tag{6}$$

The authors show that, under the assumption of homoscedasticity, the statistic

$$M_1(k) = \frac{\text{VR}(k) - 1}{\varphi(k)^{1/2}} \tag{7}$$

is asymptotically distributed as $N(0,1)$, where

$$\varphi_o(k) = \frac{2(2k-1)(k-1)}{3kT}. \tag{8}$$

To accommodate y_t's exhibiting conditional heteroscedasticity, the authors propose a second statistic, which is robust under heteroscedasticity, and follows the standard normal distribution asymptotically,

$$M_2(k) = \frac{\text{VR}(k) - 1}{\varphi^*(k)^{1/2}} \tag{9}$$

where

$$\varphi^*(k) = \sum_{j=1}^{k-1} \left[\frac{2(k-j)}{k} \right]^2 \delta(j) \tag{10}$$

$$\delta(j) = \frac{\sum_{t=j+1}^{k-1} (y_t - \hat{\mu})^2 (y_{t-j} - \hat{\mu})^2}{[\sum_{t=1}^{T} (y_t - \hat{\mu})^2]^2}. \tag{11}$$

The procedure proposed by Lo and Mackinlay (1988) is devised to test individual VR tests for a specific k-difference, but under the random walk hypothesis, we must have $\text{VR}(k) = 1$ for all k. As the null hypothesis is rejected if it is rejected for any k value, this implies that a sequential procedure of testing several k values leads to an oversized testing strategy. To account for this, Chow and Denning (1993) propose a multiple VR test where only the maximum absolute value of $\text{VR}(k)$ in a set of m test statistics is considered. The Chow–Denning test statistic is defined as

$$\text{CD}_1 = \sqrt{T} \max_{1 \leq i \leq m} |M_1(k_i)| \tag{12}$$

and it follows the studentized maximum modulus (SMM) distribution with m and T degrees of freedom, i.e., $\text{SMM}(\alpha, m, T)$. The null hypothesis (random walk) is rejected at α level of

significance if the MV_1 statistic is greater than the $[1 - (\alpha^*/2)]$th percentile of the standard normal distribution, where $\alpha^* = 1 - (1 - \alpha)^{1/m}$. However, this statistic is only valid under homoscedastic returns.

The heteroscedastic-robust version of the Chow–Denning test can be written as

$$CD_2 = \sqrt{T} \max_{1 \leq i \leq m} |M_2(k_i)| \tag{13}$$

and it has the same critical values as CD_1. If this test statistic exceeds the critical value at a predetermined significance level, then the martingale hypothesis is rejected.

One of the difficulties of using M_1, M_2, CD_1 and CD_2 tests is that they are based on asymptotic theory, and so statistical inference can be misleading in small samples (Richardson and Stock 1989). The use of long-time horizons for the calculation of returns reduces the number of observations and limits the value of asymptotic distributions, which are derived under the assumption that the sample size increases to infinity. To overcome the problem of small samples, the bootstrap method is a valid alternative inference tool to the asymptotic distribution of the above test statistics.

Kim (2006) proposes a variance-ratio test based on wild bootstrap, which is a resampling method that approximates the sampling distribution of the test statistic, that can be applied do data with unknown forms of conditional and unconditional heteroscedasticity (Davidson and Flachaire 2008). For example, the wild bootstrap test based on CD_2 (a joint variance-ratio test) can be conducted in three stages as below

(1) Form a bootstrap sample of T observations $y_t^* = \eta_t y_t (t = 1, \ldots, T)$, where η_t is a random sequence with $E(\eta_t) = 0$ and $E(\eta_t^2) = 1$.
(2) Calculate CD_2^*, which is the CD_2 statistic in Equation (13) from the bootstrap sample generated in stage (1).
(3) Repeat (1) and (2) sufficiently many, say n, times to form a bootstrap distribution $\{CD_2^*(j)\}_{j=1}^n$ of the test statistic.

The bootstrap distribution $\{CD_2^*(j)\}_{j=1}^n$ is used to approximate the sampling distribution of the CD_2 statistic. The p-value of the test is estimated as the proportion of $\{CD_2^*(j)\}_{j=1}^n$ greater than the CD_2 statistic computed from the original data. Following Kim (2006), we use the standard normal distribution for η_t to implement the wild bootstrap test, as other alternatives provide qualitatively similar sample results. We apply wild bootstrapping to M_2 and CD_2.

2.2.2 VR tests based on ranks and signs

Noting that the Lo–Mackinlay tests are biased and right-skewed in finite samples, Wright (2000) proposes a non-parametric alternative to conventional asymptotic VR tests, using ranks and signs. Given a sample of log returns $\{y_t\}_{t=1}^T$, let $r(y)$ be the rank of y_t among $(y_1, \ldots, y_T)$ which, under the hypothesis that y_t *is* i.i.d., is just a random permutation of the numbers $1, 2, \ldots, T$, each with equal probability. Define the rank-based VR tests R_1 and R_2 as (for $i = 1$ or 2):

$$R_i(k) = \left(\frac{(Tk)^{-1} \sum_{t=k}^T (r_{it} + \cdots + r_{it-k+1})^2}{T^{-1} \sum_{t=1}^T r_{it}^2} - 1 \right) \left(\frac{2(2k-1)(k-1)}{3kT} \right)^{-1/2} \tag{14}$$

where,

$$r_{1t} = \frac{(r(y_t) - (T + 1/2))}{\sqrt{((T - 1)(T + 1))/12}} \tag{15}$$

$$r_{2t} = \frac{\Phi^{-1}r(y_t)}{T + 1} \tag{16}$$

Φ^{-1} is the inverse of the standard normal cumulative distribution function. The test based on the signs of the first differences is given by

$$S_1(k) = \left(\frac{(Tk)^{-1} \sum_{t=k}^{T} (s_t + \cdots + s_{t-k+1})^2}{T^{-1} \sum_{t=1}^{T} s_t^2} - 1 \right) \left(\frac{2(2k - 1)(k - 1)}{3kT} \right)^{-1/2}. \tag{17}$$

where $s_t = 2u(y_t, 0)$ and $u(y_t, 0)$ is $\frac{1}{2}$ if yt is positive and $-\frac{1}{2}$ otherwise. Under the assumption that y_t is generated from a martingale difference sequence with no drift, s_t is an i.i.d. sequence with zero mean and unit variance, taking the values 1 and -1 with equal probability of 0.5, and the critical values of the test can be obtained by simulating its sampling distribution.

The Wright tests can be more powerful than the Lo–Mackinlay tests, having several attractive features: (i) they have high power against a wide range of models displaying serial correlation; (ii) they are exact under i.i.d.; (iii) the test based on signs is exact even under conditional heteroscedasticity, and (iv) the ranks-based tests display low-size distortion, even under conditional heteroscedasticity.

However, the use of several k values in the Wright tests would lead to an over rejection of the null hypothesis, as in the Lo–Mackinlay tests context (Belaire-Franch and Opong 2005). One solution to this problem is to construct a joint variance-ratio test for ranks (or signs) as was proposed independently by Belaire-Franch and Contreras (2004) and Kim and Shamsuddin (2008a, 2008b). For example, to test for the joint null hypothesis that $VR(k_i) = 1$ for $i = 1, \ldots, m$ against the alternative hypothesis that $VR(k_i) \neq 1$ for some i, using the sign-based test statistic, we can define

$$JS_1 = \max_{1 \leq i \leq m} |S_1(k_i)| \tag{18}$$

similarly to the Chow–Denning test. The JS_1 statistic also has an exact sampling distribution, and its critical values can be obtained by simulation. The null hypothesis is rejected when the observed JS_1 statistic is greater than its critical value.

3.   The data

The data consists of daily closing values of stock market indexes for UK, France, Germany, Spain, Greece and Portugal, chosen as representative for each of these markets. The stock market indexes are, respectively, FTSE 100, CAC 40, DAX 30, IBEX 35, ATHEX General Index and PSI 20. The source of all data is *Reuters*, and it includes observations from 1 January 1993 to 31 December 2007, during which the markets displayed wide movements, especially in the case of Greece, as shown visually in Figure 1.

We apply the empirical tests to the whole 15-year period, but also to a smaller period of 5 years, from 1 January 2003 to 31 December 2007. The testing of different periods has the advantage
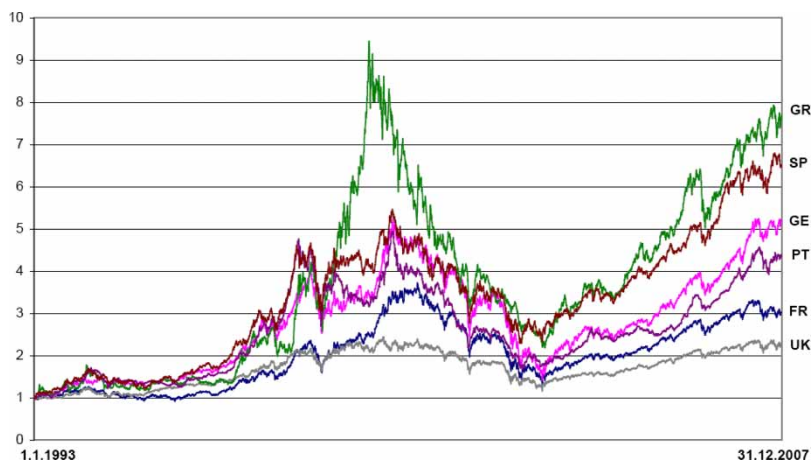
Figure 1. Stock market indexes – closing prices – 1993–2007.
Notes: This figure plots the stock market indexes of the six countries covered by the present study, between 1 January 1993 and 31 December 2007. The indexes are: FTSE 100 (for UK), CAC 40 (for France), DAX 30 (for Germany), IBEX 35 (for Spain), ATHEX General Index (for Greece) and PSI 20 (for Portugal).

of allowing for structural changes, so that EMH may be accepted in some period while in other periods that hypothesis may be rejected. We are also interested in the period from January 2003 to December 2007, because it has not been covered by previous studies, and it reflects more closely the current state of development of the markets. We deliberately leave out data for 2008, due to very large drops in prices and increased turbulence in that year, caused by the bank-originated financial and economic crisis.

Daily closing prices are used to compute weekly data. For the weekly price series, we use the observations of Wednesdays, to minimize the risk of possible weekend effects. In cases where Wednesday is missing, we use Tuesday. From country samples of around 3879 daily price observations, we generate 782 weekly price observations for 1993–2007. The returns are computed as the logarithmic difference between two consecutive prices in a series. Table 1 shows the descriptive statistics for the daily and weekly returns of the stock market indexes.

The daily returns are negatively skewed in all six countries both in 1993–2007 and in 2003–2007, which means that large negative returns tend to be larger than the higher positive returns. The level of excess kurtosis is positive for all countries, in both periods, indicating that the distributions of returns are leptokurtic, thus having higher peaks than would be expected from normal distributions. The Jarque-Bera statistic rejects the hypothesis of a normal distribution of daily returns in all countries and periods, at a significance level of 1%. An ARCH Lagrange multiplier test with 10 lags reveals that all daily returns series are strongly conditional heteroscedastic, in both periods.

In 1993–2007, the evidence from weekly returns is largely consistent with the evidence from daily returns, showing mainly negative skewness and positive excess kurtosis. Positive skewness appears for the UK in 1993–2007, and for France and Spain, in 2003–2007. Again, the Jarque-Bera statistic rejects normality of weekly returns, for all countries, in both periods. The ARCH LM test shows that conditional heteroscedasticity is present in all countries in 1993–2007, but not in Portugal and Spain, in 2003–2007.

Table 1. Descriptive statistics.

	UK		France		Germany		Spain		Greece		Portugal	
	1993–2007	2003–2007	1993–2007	2003–2007	1993–2007	2003–2007	1993–2007	2003–2007	1993–2007	2003–2007	1993–2007	2003–2007
Daily returns												
Observations	3878	1292	3878	1292	3878	1292	3878	1292	3878	1292	3878	1292
Mean return	0.0002	0.0004	0.0003	0.0005	0.0004	0.0008	0.0005	0.0007	0.0005	0.0008	0.0004	0.0006
Maximum	0.0590	0.0590	0.0700	0.0700	0.0755	0.0709	0.0632	0.0405	0.0766	0.0497	0.0694	0.0384
Minimum	−0.0589	−0.0492	−0.0768	−0.0583	−0.0887	−0.0634	−0.0734	−0.0424	−0.0962	−0.0611	−0.0959	−0.0463
St. Deviation	0.0103	0.0089	0.0130	0.0107	0.0141	0.0120	0.0126	0.0094	0.0152	0.0103	0.0097	0.0068
Skewness	−0.2145	−0.1375	−0.1068	−0.0588	−0.2685	−0.0461	−0.2104	−0.3115	−0.1102	−0.3197	−0.6442	−0.3690
Excess Kurtosis	3.1670	4.0828	2.8273	3.8577	3.6514	3.9134	3.0243	2.3142	4.3767	2.5914	8.2305	3.8434
Jarque-Bera	1650**	901.4**	1299**	801.9**	2201**	824.9**	1506**	309.2**	3103**	383.5**	11214**	824.5**
ARCH LM(9)	809.9**	277.1**	682.4**	305.5**	858.6**	323.0**	656.8**	197.2**	493.1**	105.2**	439.9**	73.3**
Weekly returns												
Observations	781	261	781	261	781	261	781	261	781	261	781	261
Mean return	0.0011	0.0019	0.0014	0.0023	0.0021	0.0038	0.0024	0.0034	0.0026	0.0041	0.0019	0.0031
Maximum	0.1359	0.1359	0.1663	0.1663	0.1715	0.1715	0.1239	0.1239	0.1447	0.0742	0.1212	0.0669
Minimum	−0.1039	−0.0808	−0.1292	−0.0997	−0.1522	−0.1257	−0.1333	−0.0755	−0.1401	−0.0988	−0.1132	−0.0614
St. Deviation	0.0221	0.0199	0.0289	0.0234	0.0307	0.0269	0.0286	0.0214	0.0362	0.0243	0.0249	0.0166
Skewness	0.0100	0.3340	−0.1380	0.4850	−0.5212	−0.0103	−0.5322	0.1322	−0.1101	−0.6645	−0.2886	−0.3879
Excess Kurtosis	3.7833	9.2710	3.4312	10.1820	3.6483	7.5161	1.9254	4.2015	1.9475	1.3870	2.8112	2.2757
Jarque-Bera	465.8**	939.6**	385.6**	1138**	468.5**	614.3**	157.5**	192.7**	1250**	40.13**	268.0**	62.87**
ARCH LM(9)	125.0**	27.0**	162.1**	52.1**	134.0**	132.4**	75.0**	30.8**	55.4**	16.6	90.0**	17.3

Notes: The Jarque-Bera test is a goodness-of-fit measure of departure from normality, based on the sample kurtosis and skewness, and is distributed as a chi-squared with two degrees of freedom. The null hypothesis is a joint hypothesis of both the skewness and excess kurtosis being 0, since samples from a normal distribution have an expected skewness of 0 and an expected excess kurtosis of 0. As the definition of *JB* shows, any deviation from this increases the *JB* statistic. The ARCH Lagrange multiplier tests the presence of ARCH effects in the return series.
**Null hypothesis rejection significant at the 1% level.

4. Results

4.1 *Runs test*

The results of the runs test, which do not depend on normality of returns, are presented in Table 2, for daily and weekly returns.

Considering the period 1993–2007, the number of runs is significantly less than expected in Portugal and Greece, both in daily and weekly returns, which is consistent with positive serial correlation of returns and is evidence against EMH, in these two countries. The other four countries display more runs than expected, although only significantly in weekly returns, in France. This mean-reverting behavior is also inconsistent with EMH, because it implies that weekly returns can be predicted from the previous weekly return.

In 2003–2007, at the 1% level of significance, the number of runs with daily data is above expected in France and Germany and below expected in Greece. With weekly data, we only find evidence of an anomalously high number of runs at the 5% level, in France and UK, which is consistent with mean reversion. In the other four countries, Germany, Spain, Greece, and Portugal, the number of runs is close to the expected.

4.2 *VR tests*

Table 3 presents the results of the VR tests which are not distorted by the presence of conditional heteroscedasticity in the returns series, i.e. CD_2, CD_2^* and JS_1. The results for M_1, M_2, CD_1, M_2^*, R_1, R_2, S_1, JR_1 and JR_2 are also computed. In order to facilitate comparisons with other recent studies, we adopt the procedure of selecting lags 2, 5, 10 and 30, for daily data, and lags 2, 4, 8, 16, for weekly data. All the variance-ratio tests have been performed using the package 'vrtest' authored by Kim (2009), which runs on R Software (R Development Core Team 2009). According to the Monte Carlo findings of Kim and Shamsuddin (2008a), the wild bootstrap (CD_2^*) and the joint sign (JS_1) tests show no size distortion, possess much higher power than the Chow–Denning test in small samples.

In 1993–2007, Greece and Portugal display VRs in daily data larger than unity, indicating that variances grow more than proportionally with time, which is consistent with the previous finding of fewer runs than expected. For these two countries, EMH is rejected at the 1% by all the VR tests, including those not reported in Table 3. An opposing different result is found for Germany and Spain, in which none of the variance-ratio tests rejects EMH. For France and UK, the joint signs variance tests clearly reject EMH, due to VRs below 1. Lo and MacKinlay (1988) show that for $k = 2$, the estimator of the VR minus one and the first-order autocorrelation coefficient estimator of weekly returns are asymptotically equal. Therefore, there is a positive serial correlation in daily returns of 0.15 in Portugal and 0.14 in Greece. In France, UK, Germany and Spain, first-order serial correlation is much closer to zero.

In 2003–2007, the heteroscedastic robust Chow–Denning statistic (CD_2) does not allow the rejection of the null hypothesis in any of the countries, even after applying wild bootstrapping to estimate confidence intervals (CD_2^*). In the opposite direction, the evidence against EMH from the joint signs-based variance test becomes stronger for France and remains significant for the UK. In the case of Greece and Portugal, the evidence of the signs-based VR test, against EMH, becomes much less stronger in the last 5 years of the sample, which is consistent with the more recent migration of these two countries from emerging to developed markets, according to Standard & Poor's/IFC. Finally, for Germany and Spain, there is virtually no evidence from VR tests, which might indicate market inefficiency in this period. Figure 2 plots the VRs of daily returns,

Table 2. Runs tests.

	UK		France		Germany		Spain		Greece		Portugal	
	1993–2007	2003–2007	1993–2007	2003–2007	1993–2007	2003–2007	1993–2007	2003–2007	1993–2007	2003–2007	1993–2007	2003–2007
Panel A: Runs in daily returns												
R	1940	680	1977	714	1999	708	1936	682	1736	595	1765	654
μ_R	1940.0	646.3	1940.0	646.3	1938.4	645.9	1939.7	645.7	1935.6	646.5	1938.6	646.4
Z	0.000	1.875	1.190	3.774**	1.947	3.461**	−0.120	2.023*	−6.427**	−2.867**	−5.580**	0.423
p-Value	1.000	0.061	0.234	0.000	0.052	0.001	0.905	0.043	0.000	0.004	0.000	0.673
Panel B: Runs in weekly returns												
R	412	146	420	147	398	141	406	135	353	120	356	131
μ_R	390.0	129.7	388.3	128.5	383.6	128.8	389.1	130.1	391.0	129.7	390.3	130.4
Z	1.584	2.046*	2.291*	2.343*	1.052	1.546	1.216	0.617	−2.727**	−1.218	−2.465*	0.078
p-Value	0.113	0.041	0.022	0.019	0.293	0.122	0.224	0.537	0.006	0.223	0.014	0.938

Notes: The runs test tests for a statistically significant difference between the expected number of runs (μ_R) *vs.* the actual number of runs (R). A run is defined as sequence of successive returns with the same sign. We define as a positive/negative return any return above/below the mean return in the period. The null hypothesis is that the successive returns follow a martingale.
*Null hypothesis rejection significant at the 5% level.
**Null hypothesis rejection significant at the 1% level.

Table 3. Variance ratio tests.

	UK		France		Germany		Spain		Greece		Portugal	
	1993–2007	2003–2007	1993–2007	2003–2007	1993–2007	2003–2007	1993–2007	2003–2007	1993–2007	2003–2007	1993–2007	2003–2007
Panel A: Variance ratio tests in daily returns(lags k = 2, 5, 10, 30)												
VR(2)	0.9938	0.8924	1.0079	0.9559	0.9840	0.9757	1.0305	0.9546	1.1390	1.0684	1.1498	1.0514
VR(4)	0.8978	0.8515	0.9483	0.8701	0.9588	0.9759	0.9891	0.9500	1.2171	1.1304	1.3150	1.1368
VR(9)	0.7941	0.7171	0.8363	0.7025	0.9084	0.8457	0.9538	0.8845	1.2553	1.1461	1.4354	1.1740
VR(30)	0.6977	0.4590	0.8143	0.5295	0.9227	0.7220	1.0321	0.7005	1.3659	1.14218	1.8056	1.3645
CD_2	2.460	2.126	2.097	2.035	1.086	1.074	1.354	1.358	5.561**	2.119	5.431**	1.832
CD_2^*	0.047*	0.081	0.107	0.108	0.584	0.690	0.406	0.466	0.000**	0.076	0.000**	0.133
JS_1	3.349**	2.659*	3.931**	4.098**	1.382	0.656	0.914	2.105	6.251**	2.326	7.472**	3.660**
Panel B: VR tests in weekly returns(lags k = 2, 4, 8, 16)												
VR(2)	0.8388	0.7625	0.8239	0.7355	0.9186	0.8883	0.9038	0.8651	1.0522	1.0666	1.0703	0.9827
VR(3)	0.7922	0.6008	0.8294	0.6313	0.9345	0.7739	0.9650	0.7710	1.1324	1.0458	1.2202	1.0157
VR(7)	0.7052	0.4619	0.8086	0.5469	0.9599	0.6515	0.9693	0.6106	1.2214	1.0358	1.4386	1.2044
VR(15)	0.6771	0.2608	0.8508	0.3467	1.0422	0.4692	0.9628	0.5099	1.2902	0.8705	1.6443	1.0545
CD_2	2.603*	1.935	2.854*	2.042	1.565	1.292	1.996	1.591	1.710	0.855	3.057**	1.050
CD_2^*	0.021*	0.091	0.014*	0.056	0.260	0.677	0.084	0.276	0.145	0.677	0.005**	0.498
JS_1	0.966	1.260	2.618*	2.159	2.998**	2.758*	2.636*	2.022	2.620*	2.758*	7.511**	5.023**

Notes: The tests based on the VR test if, in a return series, the ratio of the variance of the kth difference scaled by k to the variance of the first difference tends to equal one. VR(k) is the VR for the kth difference. CD_2 is the heteroscedasticity-robust Chow–Denning joint VR test. The critical values for the Chow–Denning test are 2.4909 (5%) and 3.0222 (1%). CD_2* are the p-values of CD_2, obtained from a wild bootstrap distribution, as proposed by Kim (2006). JS_1 is a joint test for the sign variance test of Wright (2000), as proposed by Kim and Shamsuddin (2008a, 2008b). All the variance-ratio tests have been performed using the package 'vrtest' of R Software, authored by Kim (2009).
*Null hypothesis rejection significant at the 5% level.
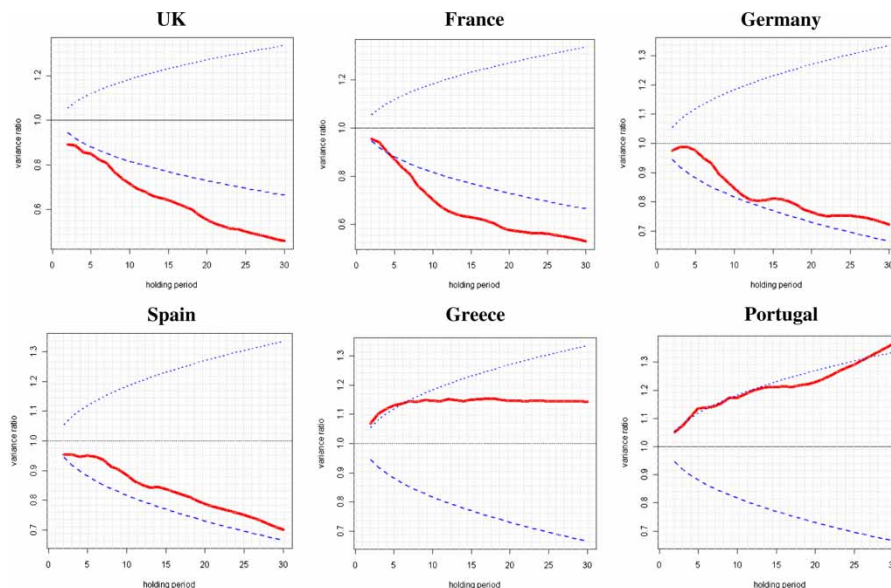**Null hypothesis rejection significant at the 1% level.

Figure 2. VRs in 2003–2007 (daily returns).
Notes: This figure shows plots of the variance ratios VR(k) for the differences $k = 2, 3, \ldots 30$, computed from daily returns of stock market indexes, in the period 2003–2007. The indexes are: FTSE 100 (for UK), CAC 40 (for France), DAX 30 (for Germany), IBEX 35 (for Spain), ATHEX General Index (for Greece) and PSI 20 (for Portugal). Also depicted is the 95% confidence band for each VR(k), using standard errors, under i.i.d. returns. All the figures have been produced using the package 'vrtest' of R Software, authored by Kim (2009).

in 2003–2007, for $k = 2$ to 30, against the 95% confidence band for each $V(k)$, using standard errors, under i.i.d. returns.

Table 3 also presents the results of the VR tests, for weekly returns. In 1993–2007, the stronger evidence against EMH from weekly returns is found in Portugal, where all the tests detect departures from market efficiency, due to VRs above unity. The estimated first-order autocorrelation coefficient for Portugal is 0.07. JS_1 is also significant for Greece and Spain, at the 5% level, for a first-order autocorrelation coefficient of 0.05. CD_2^* displays significant results, at the 5% level, for France and UK (for $k = 2$), due to variance-ratios well below unity (-0.18 for France, -0.08 for Germany, -0.16 for UK and -0.10 for Spain). JS_1 is significant at the 1% level for Germany, and at the 5% level for France.

In 2003–2007, in weekly returns, there is weak evidence from the joint-VRs tests against EMH, both in Germany and Greece (only from the signs-based tests), and none for France, UK and Spain, although the variance-ratios are much lower than unity, in this subset of the data (Figure 3). The sign-based test still rejects the null in Portugal at the 1% level.

The overall picture of the variance-ratio tests is that the highest conformity with EMH occurs in Spain and Germany, while France and UK fail at several of the tests. In fact, France and UK seem more distant from market efficiency in more recent years, due to VRs well below unity, consistently with strong mean reversion. Greece and Portugal show VRs above unity, but the level of efficiency of these markets has clearly improved in the most recent years 2003–2007.
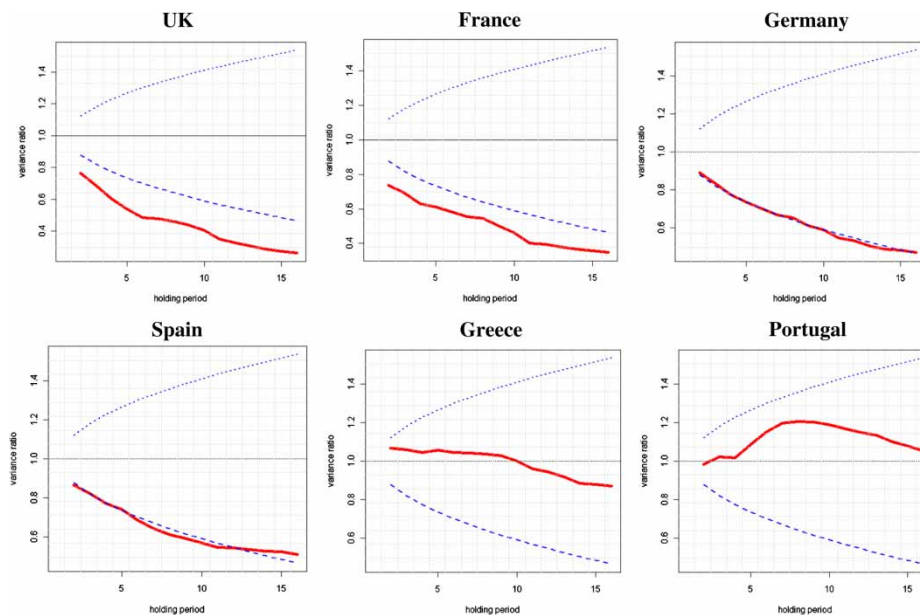
Figure 3. VRs in 2003–2007 (weekly returns).
Notes: This figure shows plots of the variance ratios VR(k) for the differences $k = 2, 3, \ldots 30$, computed from daily returns, in the period 2003–2007. The indexes are: FTSE 100 (for UK), CAC 40 (for France), DAX 30 (for Germany), IBEX 35 (for Spain), ATHEX General Index (for Greece) and PSI 20 (for Portugal). Also depicted is the 95% confidence band for each VR(k), using standard errors, under i.i.d. returns. All the figures have been produced using the package 'vrtest' of R Software, authored by Kim (2009).

5.   Comparison with previous studies

Recent studies of weak-form efficiency overlapping some of the European markets included in this study include Smith and Ryoo (2003), Worthington and Higgs (2004), Borges (Forthcoming) and Kim and Shamsuddin (2008b). Smith and Ryoo (2003) use the Chow–Denning VR test on weekly data for five European emerging markets indexes, covering 1991–1998, and reject the random walk hypothesis for Greece, Hungary, Poland and Portugal but find that Turkey follows a random walk. In the cases of Greece and Portugal, their results are similar to ours, with variance-ratios above unity, increasing with the k-differences (they use $k = 2, 4, 8, 16$), and they attribute this inefficiency to the positive serial correlation. The lower values obtained for VR(2) in the present study are probably explained by the fact that we use much more recent data, and find that the degree of market inefficiency in Greece and Portugal has reduced in recent years. Our results are also consistent with Borges (Forthcoming) who uses several different tests to show that the level of market efficiency of Portugal has increased in recent years, up to 2007.

Worthington and Higgs (2004) conduct a very detailed study of twenty European countries, from 1995 to 2003, applying multiple testing procedures, including a serial correlation test, a runs test, an augmented Dickey Fuller test and Lo–Mackinlay VR tests. They find that all indexes are not well explained by the normal distribution, and only five countries meet the most stringent criteria for a random walk, namely, Germany, Ireland, Portugal, Sweden and the UK, while France, Finland, the Netherlands, Norway and Spain meet only some of the requirements for a random walk. Overall, their results are not consistent with our results. At the 5% level, they find positive

serial correlation of lag 1 in France and UK, negative serial correlation in Greece and Spain, and no autocorrelation in Germany and Portugal, while we find the strongest evidence of daily positive autocorrelation in Portugal and Greece. Their runs tests find that Portugal and Greece have fewer runs as expected (as we do) but also that Spain has fewer runs than expected (as we do not). Also conflicting with our results is their finding of fewer runs than expected in France. Finally, in the Lo–Mackinlay heteroscedastic robust VR tests, they do not find inefficiency in any of the six countries, except in the case of Greece, for $k = 20$. Globally, Worthington and Higgs (2004) find that Portugal, Germany and UK meet the most stringent criteria for market efficiency using the daily data. This is conflict with the results of the present study, especially in the case of Portugal, where we find the strongest evidence of market inefficiency, before 2003. One major problem of the tests performed by Worthington and Higgs (2004) is that they rely on i.i.d. returns, which is clearly an invalid assumption for all six countries, as we show in Table 1. Therefore, we believe the results of the present study warrant more credibility.

In a very recent study, Kim and Shamsuddin (2008b) analyze a set of 53 international equity markets, including the six markets we cover in this study. They use daily returns from 1998 to 2007, and so their results are more comparable with the present study, considering the overlapping time horizons, and the fact that they also use the wild bootstrap Chow–Denning test (CD_2^*) and the joint sign-based VR test (JS_1). They find France, Spain and UK to be efficient, and Germany, Greece and Portugal to be inefficient. The main difference with the present study is that we reject EMH for France and UK, under the JS_1 test, and due to strong mean-reversion of returns.

Table 4. Summary of test results: EMH rejected?

	UK	France	Germany	Spain	Greece	Portugal
1993–2007: Daily returns						
Runs test	–	–	–	–	MA**	MA**
Bootstrap joint VR (CD_2^*)	–	–	–	–	MA**	MA**
Joint sign VR (JS_1)	MR**	MR**	–	–	MA**	MA**
1993–2007: Weekly returns						
Runs test	–	MR*	–	–	MA**	MA*
Bootstrap joint VR (CD_2^*)	MR*	*	–	–	–	MA**
Joint sign VR (JS_1)	–	*	MR**	MR*	MA*	MA**
2003–2007: Daily returns						
Runs test	–	MR**	MR**	MR*	MA**	–
Bootstrap joint VR (CD_2^*)	–	–	–	–	–	–
Joint sign VR (JS_1)	MR*	MR**	–	–	–	MA**
2003–2007: Weekly returns						
Runs test	MR*	MR*	–	–	–	–
Bootstrap joint VR (CD_2^*)	–	–	–	–	–	–
Joint sign VR (JS_1)	–	–	MR*	–	MA*	MA**

Notes: This table summarizes the main results of the tests. MA means that the departure from EMH is due to positive autocorrelation of returns (mean-aversion or persistence). MR means that the departure from EMH is due to negative autocorrelation of returns (mean-reversion).
*The null hypothesis of EMH is rejected at the 5% level.
**The null hypothesis of EMH is rejected at the 1% level.

6.   Conclusions

Table 4 summarizes the results of all the tests performed.

Our tests provide mixed evidence on the level of efficiency of the six stock markets analyzed. Positive first-order autocorrelation has been very strong in daily returns, in the case of Greece and Portugal, but declined in the last 5 years. This evidence of persistence in returns in these two countries is consistent with the additional findings of: (i) fewer runs than expected and (ii) VRs tend to grow with k. These two countries show clear signs of increased market efficiency, when we consider only more recent data, from 2003 to 2007, which is consistent with their migration from emerging markets to developed markets, at the turn of the twenty-first century, and according to Standard and Poor/IFC criteria.

In two of the countries where *a priori* we would expect to find the higher levels of efficiency, France and UK, we uncover the presence of strong mean reversion in weekly returns. Also, in the most recent years 2003–2007, the results show that the mean reversion has become even stronger, with first-order autocorrelation coefficients of around -0.29, for both countries. Within the group of six countries, Germany is among the most efficient, but Spain is clearly the one with higher levels of market efficiency, both in daily and weekly returns.

Two additional comments are relevant. First, the fact that different studies sometimes find contradicting evidence on EMH for the same countries and the same time periods, underlines the importance of replicating studies. Second, several studies suggest that market efficiency tends to develop over time, which justifies updating previous studies, using more recent data and a new set of more powerful techniques, based on joint variance-ratio tests, only developed in recent years.

References

Abraham, A., F. Seyyed, and S. Alsakran. 2002. Testing the random behavior and efficiency of the Gulf stock markets. *The Financial Review* 37, no. 3: 469–80.

Al-Khazali, O., D. Ding, and C. Pyun. 2007. A new variance ratio test of random walk in emerging markets: A revisit. *The Financial Review* 42, no. 2: 303–7.

Belaire-Franch, J., and D. Contreras. 2004. Ranks and signs-based multiple variance ratio tests. Working Paper, Department of Economic Analysis, University of Valencia.

Belaire-Franch, J., and K. Opong. 2005. A variance ratio test of the behaviour of some FTSE equity indices using ranks and signs. *Review of Quantitative Finance and Accounting* 24, no. 1: 93–107.

Borges, M. Forthcoming. Random walk tests for the Lisbon stock market. *Applied Economics*, DOI: 10.1080/00036840802584935 (first published on-line on 17 February 2009).

Charles, A., and O. Darné. 2009. Variance-ratio tests of random walk: An overview. *Journal of Economic Surveys* 23, no. 3: 503–27.

Chow, K., and K. Denning. 1993. A simple multiple variance ratio test. *Journal of Econometrics* 58, no. 3: 385–401.

Davidson, R., and E. Flachaire. 2008. The wild bootstrap: Tamed at last. *Journal of Econometrics* 146, no. 1: 162–9.

Fama, E. 1970. Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25, no. 2: 283–417.

Fama, E., and K. French. 1988. Permanent and temporary components of stock prices. *Journal of Political Economy* 96, no. 2: 246–73.

Grieb, T., and M. Reyes. 1999. Random walk tests for Latin American equity indexes and individual firms. *Journal of Financial Research* 22, no. 4: 371–83.

Groenewold, N., and M. Ariff. 1998. The effects of de-regulation on share market efficiency in the Asia-Pacific. *International Economic Journal* 12, no. 4: 23–47.

Huang, B. 1995. Do Asian stock markets follow random walks? Evidence from the variance ratio test. *Applied Financial Economics* 5, no. 4: 251–6.

Kim, J. 2006. Wild bootstrapping variance ratio tests. *Economics Letters* 92, no. 1: 38–43.

Kim, J. 2009. vrtest: Variance ratio tests and spectral shape tests for weak-form market efficiency. R package version 0.94. http://CRAN.R-project.org/package=vrtest (accessed July 25, 2009).

Kim, J., and A. Shamsuddin. 2008a. Are Asian stock market efficient? Evidence from the new multiple variance ratio tests. *Journal of Empirical Finance* 15, no. 3: 518–32.

Kim, J., and A. Shamsuddin. 2008b. What drives international equity market efficiency? Working Paper. http://programs.business.utsa.edu/swfa2009/PDF/102.pdf (accessed July 12, 2009).

Lo, A., and A. Mackinlay. 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies* 1, no. 1: 41–66.

Magnusson, M., and B. Wydick. 2002. How efficient are Africa's emerging stock markets. *Journal of Development Studies* 38, no. 4: 141–56.

R Development Core Team. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL. http://www.R-project.org (accessed July 25, 2009).

Richardson, M., and J. Stock. 1989. Drawing inferences from statistics based on multiyear asset returns. *Journal of Financial Economics* 25, no. 2: 323–48.

Smith, G. 2009. Martingales in European emerging stock markets: Size, liquidity and market quality. *The European Journal of Finance* 15, no. 3: 249–62.

Smith, G., K. Jefferis, and H. Ryoo. 2002. African stock markets: Multiple variance ratio tests of random walks. *Applied Financial Economics* 12, no. 7: 475–84.

Smith, G., and H. Ryoo. 2003. Variance ratio tests of the random walk hypothesis for European emerging stock markets. *The European Journal of Finance* 9, no. 3: 290–300.

Urrutia, J. 1995. Tests of random walk and market efficiency for Latin American emerging markets. *Journal of Financial Research* 18, no. 3: 299–309.

Worthington, A., and H. Higgs. 2004. Random walks and market efficiency in European equity markets. *Global Journal of Finance and Economics* 1, no. 1: 59–78.

Wright, J. 2000. Alternative variance-ratio tests using ranks and signs. *Journal of Business & Economic Statistics* 18, no. 1: 1–9.