



Article

Mean Reversions in Major Developed Stock Markets: Recent Evidence from Unit Root, Spectral and Abnormal Return Studies

James Nguyen ^{1,*}, Wei-Xuan Li ² and Clara Chia-Sheng Chen ³

¹ College of Business, Engineering and Technology, Texas A&M University Texarkana, Texarkana, TX 75503, USA

² School of Business, Stockton University, Galloway, NJ 08205, USA; liw1@stockton.edu

³ School of Business, Madonna University, Livonia, MI 48150, USA; ccchen@madonna.edu

* Correspondence: james.nguyen@tamut.edu

Abstract: We revisited the issue of return predictability in three major developed markets (USA, UK and Japan) using a unique dataset from the Wharton Research Data Services database and a comprehensive set of traditional and recent statistical methods. We specifically employed a variety of traditional linear and nonlinear tests, latest multiple-break unit root tests and spectral analysis to test the efficient market hypothesis. Our results show that these stock markets generally are inefficient. We further explored whether the departure from market efficiency can be used to generate profitable trades and found that abnormal returns exist in all three markets. We found evidence of abnormal returns associated with the break dates identified in the models which are correlated with major historical events around the world. Our findings have important implications for investors and policymakers.



Citation: Nguyen, James, Wei-Xuan Li, and Clara Chia-Sheng Chen. 2022. Mean Reversions in Major Developed Stock Markets: Recent Evidence from Unit Root, Spectral and Abnormal Return Studies. *Journal of Risk and Financial Management* 15: 162. <https://doi.org/10.3390/jrfm15040162>

Academic Editors: James W. Kolari and Seppo Pynnonen

Received: 19 January 2022

Accepted: 24 March 2022

Published: 1 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: efficient market hypothesis; unit root; spectral analysis; abnormal returns

1. Introduction

The efficient market hypothesis (EMH), introduced by Eugene Fama in 1970, states that financial asset prices entirely reflect all available information, making it impossible for investors to beat the market. The EMH posits that stock prices are sensitive to every bit of information in the market and that movements of stock prices are unpredictable. Therefore, there should not be a momentous difference between the optimal forecast and actual stock prices, and the probability of making abnormal profits in the stock market is asymptotically zero. The theory has attracted many supporters as well as critics. Shiller (1981) documented that stock price variation should not be explained by fundamentals. Some of the results which show little alpha (risk-adjusted return) and no persistence were published by Carhart (1997), Lettau and Van Nieuwerburgh (2008), Fama and French (2010), Busse et al. (2010), Bertone et al. (2015), etc. Richard Thaler, a Nobel laureate in Economics in 2017, has helped reignite this debate. Thaler, one of the founders of “behavioral finance”, has put the notion of the EMH in doubt and provided scientific explanations for the existence of irrational market behaviors. The empirical evidence is mixed, and the research community is “torn” between the EMH and behavioral finance camps (Verheyden et al. 2015).

A review of the EMH in developed markets reveals a widespread but not definitive consensus that markets tend toward efficiency, although there are periods of informational inefficiency and periods of speculative bubbles (behavioral finance) (e.g., French and Roll 1986; De Long et al. 1990). Carhart (1997) showed that the performance of mutual funds does not reflect superior stock-picking skills. Fama and French (2010) showed that few mutual funds produce returns sufficient to cover their costs. Busse et al. (2010) found that an investment manager’s superior risk-adjusted returns are indistinguishable from zero.

Finally, Bertone et al. (2015) showed that the US market had become significantly more efficient even during very short-term intervals. More recently, Durusu-Ciftci et al. (2017) argued that the evidence for the EMH is mixed. One reason is that traditional tests ignore the presence of structural breaks, leading to invalid statistical inferences. Another potential issue is that traditional unit root tests only allow for one of two breaks in the data—a problem that can be overcome by some of the multiple-break unit root tests employed in our study.

Our research contributes to the literature by testing market efficiency in three major developed markets, the USA, the UK and Japan, for the first time—to our knowledge—using unique authoritative stock price indices provided by the WRDS. Our study also complements those that examine this topic for major stock markets, especially the study of the US, UK and Japanese stock markets by Urquhart and McGroarty (2016), Urquhart and Hudson (2013), Borges (2010) and Narayan and Smyth (2007). However, we employed a number of recent and powerful statistical tests to study this issue. Specifically, in this paper, we utilized highly regarded tests such as those used by Elliott et al. (1996), Ng and Perron (2001) and Brock et al. (1996, BDS) which had not been widely used in this line of research in addition to the highly popular traditional statistical tests such as the BDS and variance ratios. Further, we took advantage of the latest multiple-break unit root tests by Lumsdaine and Papell (1997, LP), Lee and Strazicich (2003, LS), Narayan and Popp (2010, NP) and Ender and Lee (2012, EL).¹ To increase the robustness of our results, we adopted recent spectral tests commonly found in the electrical engineering literature to further assess the EMH in the three developed markets in question. The final novelty of our study is the analysis of abnormal returns. Specifically, we explored whether the departure from market efficiency can be used to generate profitable trades.

By way of preview, we found that the three stock market indices in our study exhibit mean reversions. The rather surprising finding of market inefficiency (contradicting many prior findings of market efficiency for highly developed markets) may indicate more pronounced information asymmetry, limited competition and not fully developed financial and banking systems within these countries. The paper is organized as follows. Section 2 presents a brief review of the related studies. Section 3 discusses the data and the methodology. Section 4 discusses the empirical results. Section 5 provides some discussions of the findings. Finally, Section 6 concludes the study with some remarks.

2. Brief Literature Review

Numerous studies have explored the predictability of equity returns. Early studies documented that macroeconomic and financial variables are useful predictors of equity returns. For example, Fama and Schwert (1977) found a positive relationship between inflation and expected returns. Chen et al. (1986) showed that term spread, expected and unexpected inflation, industrial production and credit spread can explain the variations of equity returns in the US dividend yields (or dividend/price ratios) and also demonstrate the strong predictive power of equity returns (e.g., Shiller 1982; Bekaert and Hodrick 1992; Campbell and Hamao 1992; Solnik 1993; Campbell and Shiller 1988; Fama and French 1988; Ang and Bekaert 2007; Golez and Koudijs 2018). Interest rates, documented by Ang and Bekaert (2007) and Rapach et al. (2013), are reliable predictors of equity returns. Size and book-to-market ratio along with the market factor, presented by Fama and French (1992, 1993), are also important variables to predict equity returns. Examining firms' fundamentals and equity prices in the USA, Bhargava (2014) found that the following variables were important predictors: earnings per share, total assets, long-term debt, dividends per share and unemployment and interest rates.

Other studies incorporate liquidity to explore its relationship with equity returns (e.g., Amihud 2002; Bekaert et al. 2007). Amihud (2002) found a positive relationship between expected returns and contemporaneous unexpected illiquidity. Bekaert et al. (2007) documented that local market liquidity is an important determinant of equity returns in emerging markets. Another line of research examines the effect of investor sentiment

on equity returns (e.g., [Baker and Wurgler 2006, 2007](#)). [Baker and Wurgler \(2006, 2007\)](#) documented a negative relationship between investor sentiment and subsequent equity returns. [Nyberg and Pönkä \(2016\)](#) documented the predictability of other equity market returns with the information from the US market.

A number of studies most related to our current research include the following studies. [Golez and Koudijs \(2018\)](#) combined the annual stock market data for the Netherlands/UK (1629–1812), the UK (1813–1870) and the USA (1871–2015) and showed that dividend yields are stationary and consistently forecast returns over both short and long horizons. [Goetzmann et al. \(2001\)](#) estimated a new index for the New York stock market between 1815 and 1925. They found little evidence for return predictability, but data limitations forced them to approximate dividends for the period before 1870. [Mitra et al. \(2017\)](#) examined the efficiency of 31 stock index series spanning 26 countries across the world. They found periods of departure from the martingale difference hypothesis among the stock index series around the world. The results are consistent with the adaptive market hypothesis whereby stock markets remain efficient most of the time but there are periods when markets become inefficient. [Urquhart and Hudson \(2013\)](#) also empirically investigated the adaptive market hypothesis for the US, UK and Japanese markets using very long-run data. Daily data were divided into five-yearly subsamples and subjected to linear and nonlinear tests to determine how the independence of stock returns had behaved over time. Their results from the linear autocorrelation, runs and variance ratio tests reveal that each market shows evidence of being an adaptive market, with returns going through periods of independence and dependence. However, results from nonlinear tests show strong dependence for every subsample in each market. [Urquhart and McGroarty \(2016\)](#) examined the adaptive market hypothesis in S&P 500, FTSE 100, NIKKEI 225 and EURO STOXX 50 by testing stock return predictability using daily data from January 1990 to May 2014. Their results show that there are periods of statistically significant return predictability, but also periods of no statistically significant predictability in stock returns. [Narayan and Smyth \(2007\)](#) showed evidence on the random walk hypothesis in G7 stock price indices using unit root tests which allow for one and two structural breaks in the trend. Evidence of mean reversion only exists for the stock price index of Japan. In short, no consensus has been reached.

3. Data and Methodologies

Our dataset was obtained from the Wharton Research Data Services (WRDS) country price index database. A major advantage of using this database is that all price series have a consistent data format. Our sample contained daily data for 23 major stock market indices in the USA, the UK and Japan. The indices were market capitalization-weighted, adjusted for stock splits and dividends. Compustat Global—Security Daily was used to construct the indices. The portfolio was rebalanced annually at the end of the last trading day of June for each country. Observations were removed if the market capitalization was not positive or if the exchange information was missing. For firms with multiple issues, the issue with the largest market capitalization was chosen. Additionally, a security had to be in the top 50% of the market capitalization of that country and traded at the stock exchanges located within the country in question. The currency of the security price had to be consistent with its ISO currency code. Lastly, only common ordinary shares were included in the indices. We extracted the country price indices for the USA, the UK and Japan in monthly frequency. In this study, our sample period for the USA spanned from January 1926 to December 2016, while the sample period for the UK and Japan started in December 1989, ending in December 2015. We employed a battery of tests typically used in the literature as well as a number of recent methods.²

Among the most important tests for market efficiency (i.e., random walk) are unit root tests. The weak-form efficient market hypothesis states that stock prices move in a random

walk fashion, or that past prices cannot be used to predict future prices. The random walk model is commonly specified as follows:³

$$y_t = \mu + y_{t-1} + \varepsilon_t$$

where y_t is the log of price or stock index return in a number of studies, μ is the drift term and ε_t is the random disturbance term. To evaluate this hypothesis, we examined the returns on the country price indices and tested for independence of their return series. To test the random walk hypothesis it is necessary to examine the existence of a unit root in a return series. More specifically, we conducted traditional, highly regarded unit root tests and more recent single- as well as multiple-break unit root tests.⁴ We first used the following highly popular unit root tests in this study: augmented Dickey–Fuller (ADF), Phillips–Perron, Elliott–Rothenberg–Stock and Ng–Perron tests. We then employed the Zivot and Andrews (1992) test as a single-break unit root test. For multiple-break unit root tests, we utilized the models developed by Lumsdaine and Papell (1997), Lee and Strazicich (2003), Narayan and Popp (2010), Ender and Lee (2012). Finally, we computed the abnormal returns for each price index using the structural break information found in those tests.

3.1. Unit Root Tests

ADF test (1981): This is our baseline test which evaluates if a series is stationary or random-walk (unit root), mainly for comparison purposes. The null hypothesis of a unit root is rejected if the test statistic is less (or more negative) than their associated critical values:

$$y_t = a_0 + \gamma_1 y_{t-1} + \theta t + \sum_{i=2}^k \beta_i y_{t-i} + \varepsilon_t$$

where y_t in our setting is the stock index return in month t . The null hypothesis is $\gamma_1 = 1$, a unit root in the return series. One problem with the ADF method is the selection of lag length (Schwert 1989). We, therefore, used Akaike’s information criterion to select the optimal lag length (to ensure that the residual was white noise) to mitigate this issue. We also performed the Phillips and Perron (1988, PP) test, a more powerful test than the ADF test (Dickey and Fuller 1981), but with better size distortions.

ERS test (1996): This is basically a modified ADF test where Elliot, Rothenberg and Stock (ERS) show that their DF-GLS test has the power function close to the point optimal test which has better power properties. This test not only provides a higher power than the ADF and PP tests, but can also distinguish persistent stationary processes from nonstationary processes. The test has the same null hypothesis as the ADF test, and its results are interpreted similarly. To our knowledge, this was the first time the ERS tests were used to examine the market efficiency hypothesis, at least for our sample of countries.

Ng–Perron test (2001): Using the procedure in the ERS test to create efficient versions of the modified PP tests of Perron and Ng (1996), Ng and Perron (2001) showed that these tests do not have the same serious size distortions as the PP tests (used in many studies reviewed in the paper) for errors with large autoregressive and moving average roots. As a result, they can give a much higher power than the PP tests. Ng and Perron constructed four test statistics which are based on the PP tests (MZ_α and MZ_t statistics), the Bhargava (1986) (MSB) statistic and the ERS point optimal statistic (MP_T). We used the modified AIC for lag selection as suggested by the authors to maximize the power. Interpretations of results for these tests are similar to those of the ADF tests discussed above. To our knowledge, this was the first time these tests were used to examine the market efficiency hypothesis for our sample of countries.

3.2. Multiple-Break Unit Root Tests

Perron (1989) showed that structural change and unit roots are intimately related, and it is important to note that conventional unit root tests (as performed in many of the reviewed studies) are biased toward a false unit root null when the data are trend-stationary with a structural break. This observation has led to the development of a large amount of literature with unit root tests that remain valid in the presence of a break. One of the novel contributions of our study is the inclusion of multiple-break unit root tests by Lumsdaine and Papell (1997, LP), Lee and Strazicich (2003, LS), Narayan and Popp (2010, NP) and Ender and Lee (2012, EL). The main limitation of a unit root test, according to Zivot and Andrews (1992), is that it allows only for one break in the data and has a lower power than the tests described below. While Perron (1989) specified an *a priori* fixed break date, the ZA tests can endogenously determine a break date from the data.

Lumsdaine and Papell (1997): Improving on ZA, the LP multiple unit root tests allow for more than one (unknown) breakpoint in either the trend, the intercept or both the trend and the intercept of the data. We used two, four and six lags for the base model as well as automatic lag selections using the AIC and the BIC.⁵ The null hypothesis is that there is a unit root in the data. Thus, if the null hypothesis is rejected, the return series is predictable, and vice versa. It should be noted that these are computationally intensive methods when two or more breaks are selected if the dataset is fairly large (more than 500).

Lee and Strazicich (2003) showed that their model outperforms that of Lumsdaine and Papell (1997, LP) in simulations and that, unlike the LP unit root test, rejection of the null unambiguously implies a stationary trend or return predictability in our case. They also showed that the power of the tests increases substantially when two or more breaks are taken into account. It is a minimum Lagrange multiplier test for testing the presence of a unit root with two structural breaks. We employed both the “Crash” model to allow for a sudden change in level but no change in the trend and the “Break” model to account for simultaneous changes in the level and the trend. The location of breakpoints is determined endogenously by conducting a grid search to locate the minimum t-statistics. We used a 10% trimming of data points at each end of the series. The critical values for the test were provided by Lee and Strazicich (2003). It is important to note that the critical values for the model with breaks in the intercept and the trend are dependent on break locations.

Narayan and Popp (2010): This has been one of the most cited tests in recent years. Narayan and Popp showed that their model outperforms those of LP and LS. Furthermore, NP possesses a more stable power and correct size. Further, the NP test accurately recognizes the break date. Since break dates are endogenously determined within the model, this test requires no prior knowledge for possible timings of structural breaks. In our study, we considered two different models, with the first model allowing two structural breaks (level) and the second model allowing two structural breaks (level and trend). The interpretation of the model is similar to those of LP and LS. To our knowledge, the NP test has not been used to study the stock market indices of our three countries.

Ender and Lee (2012): This test, also known as the Fourier unit root test, is one of the latest tests within this class. EL surpasses the aforementioned multiple-break unit root test by reducing specification errors about break dates and their forms (gradual or sharp), leading to an increase in the power of tests. The test uses trigonometric functions to capture deviations greater than the average of the dependent variable and takes into account multiple structural breaks. A major advantage of these tests is that there is no need to know *a priori* the break dates, the exact number of breaks and the form of breaks. EL utilizes a dynamic (time variant) deterministic intercept term consisting of sine and cosine functions to determine the essence of the process or whether there is a breakpoint or nonlinear trend. EL employs a specific data-generating regression model with the smallest residual sums of square at the most appropriate frequency, as well as a more precise approximation including multiple frequencies.

3.3. Spectral Analysis

A series of tests employed in this study are the recently available spectral tests commonly used in electrical engineering.⁶ We first examined the periodogram for each country to help identify the dominant periods, cyclical properties or periodicities across different frequencies (high and low) in a series. We looked for peaks or hidden periodic components in the data. If a series seems very smooth, for example, then the values of the periodogram for low frequencies will be large relative to its other values, and vice versa. For a random walk series, all sinusoids should be of similar importance, and the periodogram will vary randomly around a constant. On the other hand, if a series exhibits very pronounced spectra at higher frequencies, this may indicate that the series is driven by dynamics or transient features that frequently come and go. In this case, we would typically consider this time series as stationary (we would typically classify it as nonstationary if spectra are more prominent near zero frequency). Further, we employed Fisher's G-test to check for the proportion of intensity represented at each specific frequency to determine if the observed peak at that frequency is random or not. Particularly, this test reveals if the series in question is white noise (i.e., a stationary process) in the sense that its maximum ordinate is not significant enough. Finally, utilizing a normalized integrated spectrum, we tested the hypothesis if observations from each of series follow a white noise process.

3.4. Abnormal Returns

Another novel feature of our work is the analysis of abnormal returns. We explored in this section whether a departure from market efficiency can be used to generate profitable trades. Since the stock markets in our study were found to be inefficient, it was interesting to explore their abnormal returns. To do this, we split the sample period by the multiple structural breaks identified in these tests into subsample periods. The random walk model and a rolling 36-month estimation period were used to compute the 1-month-ahead predicted return (\hat{y}_{t+1}):

$$y_t = c + \varepsilon_t$$

$$\hat{y}_{t+1} = \hat{c} = \frac{1}{36} \sum (y_t + y_{t-1} + \dots + y_{t-35})$$

where y_t is the return in month t , c is the constant and \hat{y}_{t+1} is the predicted return in month $t + 1$.

We then subtracted the predicted return from the realized return in each month to calculate the abnormal return (AR):

$$AR_{t+1} = y_{t+1} - \hat{y}_{t+1}.$$

Summing up the monthly abnormal returns is the cumulative abnormal return (CAR) in a subsample period:

$$CAR = \sum_{t=36}^T AR_{t+1}.$$

The importance of a structural break and its impact on abnormal profits should not be overlooked as the existence of significant abnormal returns may suggest that the market in question is inefficient.

3.5. Other Tests

Variance ratio tests: This test, after [Lo and MacKinlay \(1988\)](#), has been shown to be more powerful and reliable than the ADF tests and is robust to heteroscedasticity. It is based on the notion that if a series follows a random walk process, then the variance of its q th period difference should be q times the variance of its 1-period difference. If the variance ratio test statistic is greater than 1, then the series is positively correlated. We chose two, four, eight and 16 periods as these periods are typically chosen in the literature. The variance ratio of the Lo and MacKinlay tests whether the variance ratio is equal to 1 for

a particular holding period. For each country, we presented its variance ratio, its [Chow and Denning \(1993\)](#) joint maximum z -statistic (since we chose more than one period) and its associated p -values (we did not report the individual test statistics as they are qualitatively similar). The null hypothesis of random walk is rejected if the p -value for the z -statistic is small (i.e., less than 0.05 for a 5% significance level). We noted that for a given set of test statistics, the random walk hypothesis is rejected if any one of the variance ratios is considerably dissimilar to one. The results of this test are not reported to conserve space as they are similar to those obtained using the Lo and MacKinlay tests.

BDS test (1996): This is perhaps the most popular (nonlinear) test for detecting serial dependence in time series data, after [Brock et al. \(1996\)](#). A number of studies have found evidence of the movement of asset returns. The BDS tests the null hypothesis of independent and identically distributed (IID) process against an unknown alternative. The test is estimated for different embedding dimensions (m) and distances (e). The null hypothesis of randomness is rejected if the BDS statistic exceeds 2 for a 95% confidence and 3 for a 99% confidence. For ease of interpretation, we presented results using different dimensions ($m = 2$ –6) and $e = 0.5$. The distance e was selected to make sure a certain fraction of the total number of pairs of points in the sample lie within e of each other as this approach is most invariant to the distribution of the series in question. Furthermore, we let e vary from 0.50 to 2 (the higher this value, the lower the power of the test). The results for the tests where e was higher than 0.5 are not reported as they are similar to those of the baseline case. As a further robustness test, especially when dealing with shorter series, we also chose the option of calculating bootstrapped p -values for the test statistic using various repetitions to increase the accuracy of the p -values (the results are not shown as they are qualitatively similar to those from the standard tests).

4. Empirical Results

Table 1 presents the summary statistics of the data⁷. As found in many prior studies, all the return series for the USA, the UK and Japan were not normally distributed, based on their associated Jarque–Bera statistics. We also examined the correlation matrix (results not shown) and observed that these return series are positively (and statistically significant) related,⁸ similar to those found in other developed markets in several prior studies. The rather high kurtosis numbers suggest the higher likelihood of extreme returns in the data for all the three countries. The skewness numbers indicate high volatility, with some extreme gains for the USA and losses for the UK and Japan.

Table 1. Descriptive statistics.

	USA	UK	Japan
	y_{us}	y_{uk}	y_{jp}
Mean	0.0062	0.0070	0.0013
Median	0.0091	0.0116	0.0011
Max.	0.4222	0.1129	0.1843
Min.	−0.2994	−0.1306	−0.2012
Std. dev.	0.0542	0.0405	0.0567
Skewness	0.2928	−0.4591	−0.0535
Kurtosis	12.4360	3.6832	3.8240
Jarque–Bera	4063.0700 ***	17.0834 ***	9.0045 **
p	0.0000	0.0002	0.0111
NOB	1092	313	313

Notes: This table reports the descriptive statistics of monthly returns, y , on the country price indices for the USA, the UK and Japan. The sample period for the USA spanned from January 1926 to December 2016, while the sample period for the UK and Japan started in December 1989 to December 2015. Notations ** and *** indicate 5% and 1% significance levels, respectively.

Table 2 shows the results for simple unit root tests. At the 1% level of significance, the ADF and Phillip–Perron tests unanimously rejected the random walk hypothesis. Table 3

displays the results for the ERS tests which also rejected the null hypothesis. It is interesting to note that the random walk hypothesis was rejected by only two of the four tests for the USA, the UK and Japan indicated in Table 4 (Ng–Perron). Table 5 reports the Zivot–Andrews test results. Again, the unit root or the random walk hypothesis was rejected at the 1% significance level. These results are in line with some of the prior reviewed studies but are in stark contrast to those obtained by Narayan and Smyth (2007), except for Japan whose price series was found to be stationary. It is possible that their tests (LS, LP, Perron, Zivot and Andrews) suffer from the same problems as those discussed by Narayan and Popp (2010) and Ender and Lee (2012) which are performed in our study. While the test found a structural break in April 2000 for the US, in March 2009 for the UK and in March 2007 for Japan, these results should be interpreted with extreme caution (Perron 1989).⁹

Table 2. Unit root tests: Augmented Dickey–Fuller and Phillip–Perron Tests.

	USA		UK		Japan	
	ADF Test	Phillips–Perron Test	ADF Test	Phillips–Perron Test	ADF Test	Phillips–Perron Test
a_0	0.0046 (0.1585)	0.0046 (0.1585)	0.008795 * (0.0585)	0.008795 * (0.0585)	−0.0077 (0.2334)	−0.0077 (0.2334)
γ_1	−0.9163 *** (0.0000)	−0.9163 *** (0.0000)	−0.9444 *** (0.0000)	−0.9444 *** (0.0000)	−0.9034 *** (0.0000)	−0.9034 *** (0.0000)
θ	0.0000 (0.7117)	0.0000 (0.7117)	−0.0000 (0.5508)	−0.0000 (0.5508)	0.0000 (0.1158)	0.0000 (0.1158)
Adj. R^2	0.4572	0.4572	0.4696	0.4696	0.4480	0.4480
NOB	1092	1092	313	313	313	313

Notes: ADF denotes the augmented Dickey–Fuller test. For details, please see Dickey and Fuller (1979) and Phillips and Perron (1988). The model specification for the ADF and Phillips–Perron tests is: $y_t = a_0 + \gamma_1 y_{t-1} + \theta t + \sum_{i=2}^p \beta_i y_{t-i} + \varepsilon_t$, where y_t is the stock index return in month t . The null hypothesis is $\gamma_1 = 1$, a unit root in the return series. p -values are in the parentheses. Notations *, ** and *** denote 10%, 5% and 1% significance levels, respectively.

Table 3. Unit root tests: Elliott–Rothenberg–Stock test.

	USA	UKA	Japan
Elliott–Rothenberg–Stock test statistic	0.6490 ***	0.7920 ***	−15.398 ***
Test critical values: 1% level	3.9600	3.9915	−3.4712
Test critical values: 5% level	5.6200	5.6374	−2.9076
Test critical values: 10% level	6.8900	6.8770	−2.6008
NOB	1092	313	313

Notes: The equations of unit root testing by Elliott et al. (1996) are specified as follows: $y_t = d_t + U_t$, $U_t = \alpha U_{t-1} + v_t$, where y_t is the stock index return, d_t is a deterministic component, v_t is an unobserved stationary error with zero mean, and its spectral density at frequency of zero is a positive value. In the GLS-detrended series, $\tilde{y}_t \equiv y_t - \hat{\phi}' Z_t$, $\hat{\phi}$ minimizes $S(\bar{\alpha}, \phi) = (y^{\bar{\alpha}} - \phi' Z^{\bar{\alpha}})'(y^{\bar{\alpha}} - \phi' Z^{\bar{\alpha}})$, where Z_t is a set of deterministic components and $\bar{\alpha} = (1 + \frac{\varepsilon}{T})$. The null hypothesis of a unit root is $\alpha = 1$, while the alternative hypothesis is $\alpha = \bar{\alpha}$. The likelihood ratio statistic is defined as $L = S(\bar{\alpha}) - S(1)$, where $S(\bar{\alpha}) = \min_{\phi} S(\bar{\alpha}, \phi)$. The statistic of a feasible point optimal test is $P_T = [S(\bar{\alpha}) - S(1)] / S_{AR}^2$. S_{AR}^2 is the autoregressive estimate of the spectral density at zero frequency of v_t . $S_{AR}^2 = \hat{\sigma}_k^2 / (1 - \hat{\beta}(1))^2$. In an augmented Dickey–Fuller equation, $y_t = d_t + \gamma_1 y_{t-1} + \sum_{i=2}^k \beta_i y_{t-i} + \varepsilon_{tk}$, $\hat{\beta}(1) = \sum_{i=2}^k \hat{\beta}_i$ and $\hat{\sigma}_k^2 = (T - k)^{-1} \sum_{t=k+1}^T \hat{\varepsilon}_{tk}^2$, where T is the total of time periods and k is the lag length. Notation *** denotes a 1% significance level.

Table 4. Unit root tests: Ng–Perron test.

	USA				UK				Japan			
	MZ_α	MZ_t	MSB	MP_T	MZ_α	MZ_t	MSB	MP_T	MZ_α	MZ_t	MSB	MP_T
Ng–Perron test statistics	−96.1995 ^a	−6.9339 ^a	0.0721	0.9531	−8.1361	−2.0149	0.2476 ^a	11.2068 ^a	−153.1710 ^a	−8.7497 ^a	0.0571	0.6000
Asym. critical values: 1% level	−23.8000	−3.4200	0.1430	4.0300	−23.8000	−3.4200	0.1430	4.0300	−23.8000	−3.4200	0.1430	4.0300
Asym. critical values: 5% level	−17.3000	−2.9100	0.1680	5.4800	−17.3000	−2.9100	0.1680	5.4800	−17.3000	−2.9100	0.1680	5.4800
Asym. critical values: 10% level	−14.2000	−2.6200	0.1850	6.6700	−14.2000	−2.6200	0.1850	6.6700	−14.2000	−2.6200	0.1850	6.6700
NOB	1092	1092	1092	1092	313	313	313	313	313	313	313	313

Notes: The equations of unit root testing by [Ng and Perron \(2001\)](#) are specified as follows: $y_t = d_t + U_t$, $U_t = \alpha U_{t-1} + v_t$, where y_t is the stock index return, d_t is a deterministic component, v_t is an unobserved stationary error with zero mean, and its spectral density at zero frequency is a positive value. $d_t = \sum_{i=0}^p \phi^i t^i$. The analysis by [Ng and Perron \(2001\)](#) focused on $p = 0, 1$, but it remains valid in general cases. The null hypothesis of a unit root is $\alpha = 1$, while the alternative hypothesis is $\alpha < 1$. In an augmented Dickey–Fuller equation, $y_t = d_t + \gamma_1 y_{t-1} + \sum_{i=2}^k \beta_i y_{t-i} + \varepsilon_{tk}$, $\hat{\beta}(1) = \sum_{i=2}^k \hat{\beta}_i$ and $\hat{\sigma}_k^2 = (T - k)^{-1} \sum_{t=k+1}^T \hat{\varepsilon}_{tk}^2$. $S_{AR}^2 = \hat{\sigma}_k^2 / (1 - \hat{\beta}(1))^2$. $MZ_\alpha = (T^{-1} y_T^2 - S_{AR}^2) \left(2T^{-2} \sum_{t=1}^T y_{t-1}^2 \right)^{-1}$, $MSB = \left(T^{-2} \sum_{t=1}^T y_{t-1}^2 / S_{AR}^2 \right)^{(1/2)}$. $MZ_t = MZ_\alpha \times MSB$. The statistic for the modified feasible point optimal test by [Ng and Perron \(2001\)](#) is as follows: when $p = 0$, $MP_T^{GLS} = \left[c^{-2} T^{-2} \sum_{t=1}^T \tilde{y}_{t-1}^2 - \bar{c} T^{-1} \tilde{y}_T^2 \right] / S_{AR}^2$. When $p = 1$, $MP_T^{GLS} = \left[c^{-2} T^{-2} \sum_{t=1}^T \tilde{y}_{t-1}^2 + (1 - \bar{c}) T^{-1} \tilde{y}_T^2 \right] / S_{AR}^2$. Notation ^a denotes a 1% significance level.

Table 5. Single-break unit root tests: Zivot–Andrews test.

	USA	UK	Japan
Zivot–Andrews test statistic	−14.17325 ***	−16.8966 ***	−16.26222 ***
1% critical value	−5.57	−5.34	−5.57
5% critical value	−5.08	−4.93	−5.08
10% critical value	−4.82	−4.58	−4.82
Breakpoint	April 2000	March 2009	March 2007
NOB	1092	313	313

Notes: Zivot and Andrews (1992) modified three models developed by Perron (1989), the crash model (model A), the changing growth model (model B) and the changes in the level and slope of the trend function (model C), to endogenously determine a breakpoint from the data. The following are the modified models: Model A: $y_t = \hat{\mu}^A + \theta^A DU_t(\hat{\lambda}) + \hat{\beta}^A t + \hat{\alpha}^A y_{t-1} + \sum_{j=2}^k \hat{c}_j^A y_{t-j} + \hat{\varepsilon}_t$, model B: $y_t = \hat{\mu}^B + \hat{\beta}^B t + \hat{\gamma}^B DT_t^*(\hat{\lambda}) + \hat{\alpha}^B y_{t-1} + \sum_{j=2}^k \hat{c}_j^B y_{t-j} + \hat{\varepsilon}_t$, model C: $y_t = \hat{\mu}^C + \hat{\theta}^C DU_t(\hat{\lambda}) + \hat{\beta}^C t + \hat{\gamma}^C DT_t^*(\hat{\lambda}) + \hat{\alpha}^C y_{t-1} + \sum_{j=2}^k \hat{c}_j^C y_{t-j} + \hat{\varepsilon}_t$, where y_t in our setting is the stock index return in month t , $\lambda = T_B/T$, T_B is the breakpoint, T is the total of time periods, $DU_t(\lambda) = 1$ if $t > T\lambda$ and zero otherwise and $DT_t^*(\lambda) = t - T\lambda$ if $t > T\lambda$ and zero otherwise. Notation $\hat{\lambda}$ is the estimated value of the break function. The null hypothesis of a unit root is $\alpha = 1$. The test statistic is $t_{\hat{\alpha}^i}(\lambda)$, and $i = A, B, C$. λ was chosen to minimize the one-sided t-statistic for testing the unit root (i.e., $\alpha^i = 1$). Notation *** denotes a 1% significance level.

Table 6 presents the results of the LP multiple-break unit tests with two lags and two breaks as typically suggested in the econometric literature. First, the null hypothesis of a unit root (with two or more breaks) was rejected by both tests at the 1% significance level for the USA, the UK and Japan.¹⁰ Table 9 reports the findings for Narayan and Popp. Again, the unit root or the random walk hypothesis was rejected at the 1% significance level. The test also found two breaks. Similarly, the LS test rejected the random walk hypothesis, as shown in Tables 7 and 8. It is interesting to note that LS only found one break for the US and two breaks for the UK and Japan and that the break dates in LS and LP are quite different—a well-documented phenomenon in the literature. The NP test (Table 9) appears to do a better job in capturing breaks in all the three series which occurred around the financial crisis starting in 2007. The NP results also unambiguously rejected the null hypothesis, based on model 1 (break in the level, not reported) and model 2 allowing for breaks in both the level and the trend (shown in the table). These break dates found in the LP and NP tests were later used in the final part of our paper to study the associated abnormal returns in these countries. Finally, the findings for EL presented in Table 10 are similar to those of LP, LS and NP. Note that EL, while allowing for an unknown number of breaks, does not report the number of breaks. The optimal lags chosen to minimize the residual sum of squares were six, seven and two for the US, the UK and Japan, respectively.

Table 6. Multiple-break unit root tests: Lumsdaine–Papell test.

	USA	UK	Japan
μ	0.0037 (0.7498)	0.0161 (2.4993)	−0.0204 (−1.9811)
β	0.0000 (0.6149)	−0.0001 (−1.8426)	0.0003 (2.2255)
θ	−0.0108 (−1.4678)	0.0395 (3.4249)	−0.0505 (−2.992)
γ	0.0000 (1.4546)	−0.0007 (−3.0054)	0.0006 (1.7457)
ω	−0.025 (−2.6197)	0.0499 (3.8389)	−0.0508 (−3.0221)
ψ	0.0000 (0.5755)	0.0005 (1.7308)	−0.0006 (−1.7839)
α	−1.0284 *** (−14.094)	−1.0022 *** (17.655)	−0.9501 *** (−16.6473)
NOB	1092	313	313
Number of breaks	2	2	2
First break	March 1968	March 2003	February 2000
Second break	April 2000	February 2009	January 2006

Notes: The model specification for the [Lumsdaine and Papell \(1997\)](#) test is as follows: $y_t = \mu + \beta t + \theta DU1_t + \gamma DT1_t + \omega DU2_t + \psi DT2_t + \alpha y_{t-1} + \sum_{i=2}^k c y_{t-i} + \varepsilon_t$, where y_t in our setting is the stock index return in month t , $DU1_t$ ($DU2_t$) is an indicator dummy for a mean shift at $TB1$ ($TB2$), the time breakpoint, and $DT1$ ($DT2$) is the corresponding trend shift variable. The null hypothesis is $\alpha = 1$, a unit root in the return series. Given that $\delta_1 = TB1/T$ and $\delta_2 = TB2/T$, the test statistic is defined as $\hat{t}(\delta_1, \delta_2) \Rightarrow \int_0^1 w^*(s)dw(s) / \left[\int_0^1 w^*(s)^2 ds \right]^{(1/2)}$, where $w(s)$ is a Wiener process. T-statistics are in brackets. Notation *** denotes a 1% significance level.

Table 7. Multiple-break unit root tests: Lee and Strazicich test: the crash model.

	USA	UK	Japan
μ	−0.0120 *** (−6.1079)	0.0010 (0.4163)	0.0023 (0.7139)
δ_{DU1}	−0.0445 (−0.8311)	0.0655 (1.5284)	0.1313 (2.2769)
δ_{DU2}		0.0630 (1.4687)	0.0629 (1.1020)
ϕ	−0.9113 *** (−10.9841)	−0.5163 *** (−5.9937)	−0.7294 *** (−8.4737)
Minimum test stat. (tau)	−10.9841	−5.9937	−8.4737
Test critical values: 1% level	−3.7980	−4.2264	−4.2264
Test critical values: 5% level	−3.2300	−3.6356	−3.6356
Test critical values: 10% level	−2.9250	−3.2995	−3.2995
Breakpoint	June 1981	September 2003 February 2010	March 1993 April 2003
NOB	1092	313	313

Notes: The specification for the crash model in the [Lee and Strazicich \(2003\)](#) test is as follows: $y_t = \delta' Z_t + \phi \tilde{S}_{t-1} + \mu_t$, $\tilde{S}_t = y_t - \tilde{\psi}_x - Z_t \tilde{\delta}$, $\tilde{\psi}_x = y_1 - Z_1 \tilde{\delta}$, where Z_t is a set of exogenous variables, $Z'_t = [1, t, DU1_t, DU2_t]$ and δ' is a set of coefficients $[\delta_1, \delta_1, \delta_{DU1}, \delta_{DU2}]$. The null hypothesis is $\phi = 1$, a unit root in the return series. T-statistics are in brackets. Notation *** denotes a 1% significance level.

Table 8. Multiple-break unit root tests: Lee and Strazicich test: the break model.

	USA	UK	Japan
μ	−0.0161 *** (−7.0990)	−0.0508 *** (−11.2039)	−0.0339 *** (−6.5624)
δ_{DU1}	−0.0192 (−0.3569)	−0.2457 *** (−5.9904)	0.0995 (1.7233)
δ_{DU2}		−0.0983 (−2.5119)	−0.1006 (−1.7423)
δ_{DT1}	0.0124 (2.4237)	0.1138 (8.4136)	−0.0481 * (−4.8467)
δ_{DT2}		−0.0273 (−2.2535)	0.0965 (7.0711)
ϕ	−0.9139 *** (−10.9961)	−1.0956 *** (−13.7221)	−0.8416 *** (−11.0827)
Minimum test stat. (tau)	−10.9961	−13.7221	−11.0827
Test critical values: 1% level	−4.4612	−5.6458	−5.5177
Test critical values: 5% level	−3.9240	−4.9246	−5.0260
Test critical values: 10% level	−3.6492	−4.6474	−4.7586
Breakpoint	November 2005	August 2008 September 2009	November 2005 March 2010
NOB	1092	313	313

Notes: The specification for the break model in the Lee and Strazicich test is as follows: $y_t = \delta'Z_t + \phi\tilde{S}_{t-1} + \mu_t$, $\tilde{S}_t = y_t - \tilde{\psi}_x - Z_t\tilde{\delta}$, $\tilde{\psi}_x = y_1 - Z_1\tilde{\delta}$, where Z_t is a set of exogenous variables, $Z_t' = [1, t, DU1_t, DU2_t, DT1_t, DT2_t]$ and δ' is a set of coefficients $[\delta_1, \delta_1, \delta_{DU1}, \delta_{DU2}, \delta_{DT1}, \delta_{DT2}]$. The null hypothesis is $\phi = 1$, a unit root in the return series. T-statistics are in brackets. Notations *** and * denote 1% and 10% significance levels.

Table 9. Multiple-break unit root tests: Narayan and Popp test.

	USA	UK	Japan
Narayan and Popp test statistic	12.666 ***	17.534 ***	16.397 ***
1% critical value	5.287	5.318	5.318
5% critical value	4.692	4.741	4.741
10% critical value	4.396	4.430	4.430
Breakpoint	July 2007 January 2009	June 2008 September 2008	August 2008 April 2009
NOB	1092	313	313

Notes: This table reports the test statistic of the model with a break and a trend in the paper by [Narayan and Popp \(2010\)](#). The null hypothesis is a unit root in the return series. The test is based on the following process: $y_t = d_t + U_t$, $U_t = t_{t-1} + \varepsilon_t$, $\varepsilon_t = \psi^*(L)\varepsilon_t = A^*(L)B(L)^{-1}e_t$, where y_t is the return series with a deterministic component d_t and a stochastic component U_t , e_t is iid $(0, \sigma^2)$ with $A^*(L)$ and $B(L)$ being polynomial lags of order p and q lying outside the unit circle. Model 1 in the paper by [Narayan and Popp \(2010\)](#) allows for two breaks in the level. Model 2 (shown) allows for two breaks in the level and the trend. Notation *** denotes a 1% significance level.

Table 10. Multiple-break unit root tests: Ender and Lee test.

	USA	UK	Japan
Ender and Lee test statistic	10.299 ***	4.175 **	8.144 ***
1% critical value	4.560	4.610	3.730
5% critical value	4.030	4.070	3.120
10% critical value	3.770	3.790	2.830
Chosen lag	6	7	2
Frequency	1	1	5
NOB	1092	313	313

Notes: Ender and Lee test (2012) is a modification of the DF test in which $d(t)$ or the time-dependent deterministic term is added to the test regression: $Y(t) = d(t) + \alpha Y_{t-1} + e_t$ and e_t is iid $(0, \sigma^2)$, where Y is the stock return. The unit root null hypothesis of $\alpha = 1$ is tested by approximating $d(t)$ with the following Fourier function: $d(t) = \phi_0 + \phi_{\sin} \cdot \sin(2\pi kt/T) + \phi_{\cos} \cdot \cos(2\pi kt/T) + \varepsilon_t$, where $\varepsilon_t = \alpha e_{t-1} + u_t$, k is the single frequency component and measures the amplitude and displacement of the sinusoidal component of $d(t)$, $t = 1, 2, \dots, T$. The above equation is estimated for all integer values of k which lie between the interval $[1, 5]$ and selecting the estimation which produces the lowest residual sum of squares. Notations *** and ** denote 1% and 5% significance levels, respectively.

The results from the normalized integrated spectrum tests are shown in Figure 1 (USA), Figure 2 (UK) and Figure 3 (Japan). The null of stationarity was rejected at the 5% significance level in all the three return series as the statistics fell within the two bands. Fisher's G-tests and periodograms for each country, not shown to save space, are qualitatively similar.

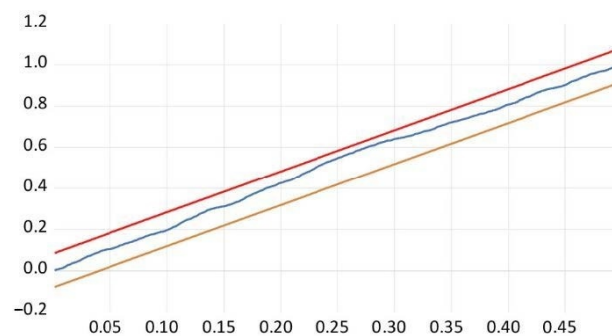


Figure 1. Spectral tests: Normalized integrated spectrum for the USA. Note: This test evaluates the null hypothesis that the data are stationary (white noise). It is based on the normalized integrated spectrum with the following statistic: $U_p = \frac{\sum_{k=1}^p I(w_q)}{\sum_{k=1}^{n/2} I(w_q)}$, where $I(\omega(i))$ is the i th maximum. The test statistic (vertical axis) is plotted against its frequency (horizontal axis). If the deviations of the statistic from the $p/(n/2)$ line do not exceed $\pm a\sqrt{2/n}$, the null will not be rejected, where a is set equal to 1.36 for 95% confidence.

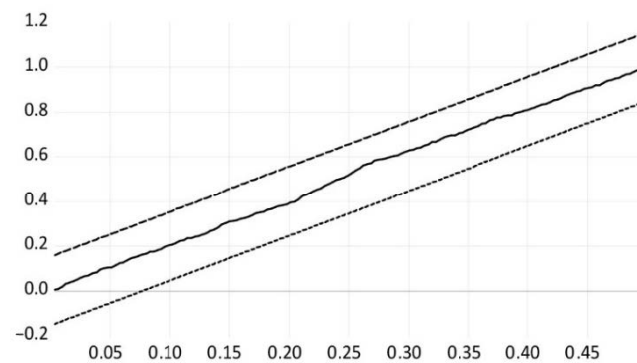


Figure 2. Spectral tests: Normalized integrated spectrum for the UK. Note: This test evaluates the null hypothesis that the data are stationary (white noise). It is based on the normalized integrated spectrum with the following statistic: $U_p = \frac{\sum_{k=1}^p I(w_k)}{\sum_{k=1}^{n/2} I(w_k)}$, where $I(\omega(i))$ is the i th maximum. The test statistic (vertical axis) is plotted against its frequency (horizontal axis). If the deviations of the statistic from the $p/(n/2)$ line do not exceed $\pm a\sqrt{2/n}$, the null will not be rejected, where a is set equal to 1.36 for 95% confidence.

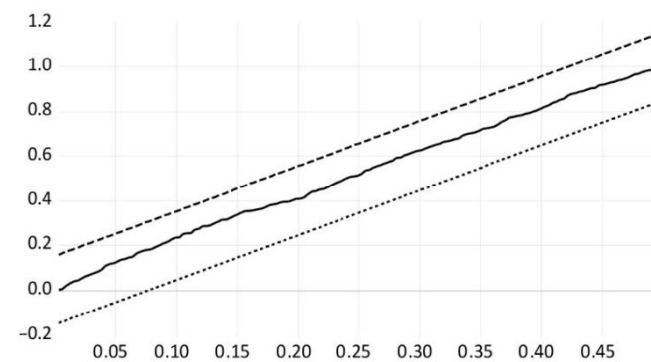


Figure 3. Spectral tests: Normalized integrated spectrum for Japan. Note: This test evaluates the null hypothesis that the data are stationary (white noise). It is based on the normalized integrated spectrum with the following statistic: $U_p = \frac{\sum_{k=1}^p I(w_k)}{\sum_{k=1}^{n/2} I(w_k)}$, where $I(\omega(i))$ is the i th maximum. The test statistic (vertical axis) is plotted against its frequency (horizontal axis). If the deviations of the statistic from the $p/(n/2)$ line do not exceed $\pm a\sqrt{2/n}$, the null will not be rejected, where a is set equal to 1.36 for 95% confidence.

Table 11 reports the mean abnormal returns and the cumulative abnormal returns for the USA, the UK and Japan. To conserve space, we presented these statistics for a sample period with two structural breaks identified by the Lumsdaine–Papell test and the Narayan and Popp test. The two structural breaks identified using the Narayan and Popp test coincide with the recent global financial crisis period. Table 11a,b show that the mean abnormal return in most of subsample periods for the USA, the UK and Japan is close to zero. However, significant cumulative abnormal returns are found for the USA, the UK and Japan, lending support for market inefficiency. Interestingly, the cumulative abnormal returns were all positive (negative) for Japan (UK) in these subsample periods. The cumulative abnormal returns in the subsample periods ranged from 14.64% to 81.02% for Japan, whereas they were between −4.05 and −64.51% for the UK. The positive (negative) cumulative abnormal returns for Japan indicated that the stock market in Japan (UK) consistently outperformed (underperformed) the random walk model.

Table 11. Abnormal and cumulative abnormal returns for the USA, the UK and Japan. (a) Sample period split by breakpoints identified by the Lumsdaine–Papell test. (b) Sample period split by breakpoints identified using the Narayan and Popp test.

(a)					
USA					
	Subsample period 1	First breakpoint	Subsample period 2	Second breakpoint	Subsample period 3
	Jan. 1926–Feb. 1968	Mar. 1968	Apr. 1968–Mar. 2000	Apr. 2000	May 2000–Dec. 2016
Mean Ab. Ret	−0.0010 (0.0671)		0.0007 (0.0447)		0.0018 (0.0405)
Cum. Ab. Ret NOB	−0.4511 470		0.2281 348		0.2924 164
UK					
	Subsample period 1	First breakpoint	Subsample period 2	Second breakpoint	Subsample period 3
	Dec. 1989–Feb. 2003	Mar. 2003	Apr. 2003–Jan. 2009	Feb. 2009	Mar. 2009–Dec. 2015
Mean Ab. Ret	−0.0042 (0.0393)		−0.0190 (0.0441)		−0.0029 (0.0313)
Cum. Ab. Ret NOB	−0.5166 123		−0.6451 34		−0.1334 46
Japan					
	Subsample period 1	First breakpoint	Subsample period 2	Second breakpoint	Subsample period 3
	Dec. 1989–Jan. 2000	Feb. 2000	Mar. 2000–Dec. 2005	Jan. 2006	Feb. 2006–Dec. 2015
Mean Ab. Ret	0.0077 (0.0572)		0.0238 (0.0407)		0.0093 (0.0508)
Cum. Ab. Ret NOB	0.6635 86		0.8102 34		0.7698 83
(b)					
US					
	Subsample period 1	First breakpoint	Subsample period 2	Second breakpoint	Subsample period 3
	Jan. 1926–Jun. 2007	Jul. 2007	Aug. 2007–Dec. 2008	Jan. 2009	Feb. 2009–Dec. 2016
Mean Ab. Ret	−0.0003 (0.0567)		N/A N/A		−0.0014 (0.0304)
Cum. Ab. Ret NOB	−0.2935 942		N/A N/A		−0.0800 59
UK					
	Subsample period 1	First breakpoint	Subsample period 2	Second breakpoint	Subsample period 3
	Dec. 1989–May 2008	June. 2008	Jul. 2008–Aug. 2008	Sep. 2008	Oct. 2008–Dec. 2015
Mean Ab. Ret	−0.0008 (0.0368)		N/A N/A		−0.0008 (0.0321)
Cum. Ab. Ret NOB	−0.1555 186		N/A N/A		−0.040538 51
Japan					
	Subsample period 1	First breakpoint	Subsample period 2	Second breakpoint	Subsample period 3
	Dec. 1989–Jul. 2008	Aug. 2008	Sep. 2008–Mar. 2009	Apr. 2009	May 2009–Dec. 2015
Mean Ab. Ret	0.0008 (0.0525)		N/A N/A		0.0051 (0.0513)
Cum. Ab. Ret NOB	0.146402 188		N/A N/A		0.2248 44

Notes: A rolling 36-month estimation period was used to compute the 1-month-ahead predicted return from the random walk model. Each month, the predicted return is subtracted from the realized return to obtain an abnormal return. The cumulative abnormal return is the sum of abnormal returns in a subsample period. The following are the specifications of the random walk model, predicted return (\hat{y}), abnormal return (AR) and cumulative abnormal return (CAR): $y_t = c + \varepsilon_t$, $\hat{y}_{t+1} = \hat{c} = \frac{1}{36} \sum (y_t + y_{t-1} + \dots + y_{t-35})$, $AR_{t+1} = y_{t+1} - \hat{y}_{t+1}$, $CAR = \sum_{t=36}^T AR_{t+1}$. If a subsample period is shorter than 36 months, predicted return, abnormal return and cumulative abnormal return are not computed. The standard deviation of abnormal returns is in parentheses.

Table 11a also shows that the cumulative abnormal return for the USA was -45.11% , 22.81% , and 29.24% in the periods between January 1926 and February 1968, April 1968 and March 2000, May 2000 and December 2016, respectively. In Table 11b, we find cumulative abnormal returns of -29.35% and -8% for the USA during the periods of January 1926 to June 2007 and February 2009 to December 2016, respectively. The presence of significant cumulative abnormal returns again suggests that the US stock market is not efficient. Overall, our empirical evidence implies that abnormal profits can be exploited if structural breaks are correctly identified and appropriate trading strategies are implemented. The importance of a structural break and its impact on abnormal profits cannot be overlooked.

It is important to note that the structural breaks found correspond to major historical economic events. For example, 1968 is the year of economic crisis in the USA (Collins 1996): the Bretton Woods Agreement caused the balance of payments deficit in the USA. In March 1968, foreign investors started selling US dollars to buy gold, which led to the crack of the Bretton Woods Agreement. In April 2000, the NASDAQ Composite Index plummeted 10% (Johansen and Sornette 2020). When the UK joined the Iraq War in March 2003, the FTSE 100 Index hit bottom at 3272. In January 2009, the UK entered the recession, and the unemployment rate rose in February 2009. For Japan, the recession of the Japanese economy started in the 1990s and continued to 2002. The Nikkei 225 Index rose above 20,000 yen in March 2000 because of the dot.com spillover effect from the USA. In the same month, news that Japan had entered a recession led to a global selloff which adversely affected technology stocks. In January 2006, Japan continued its expansion which started in 2002.

5. Discussion

The overall findings of mean reversions in our study may suggest that stock index prices behave in an ergodic manner. Horst and Wenzelburger (2008) showed in a theoretical model that financial market dynamics is ergodic if the interaction between households is sufficiently weak. In this case, market shares settle down to a unique equilibrium. However, when ergodicity no longer holds (if the interactive complementarities in the financial market are “too powerful”), “history matters” and the long-run market shares of competing financial mediators are path-dependent.

Our results also lend support to the existence of “market anomalies” or “behavioral finance” as discussed in earlier sections of the paper. Even in an imperfectly efficient market, Grossman and Stiglitz (1980) showed that there still exist opportunities for abnormal investment returns due to superior information gathering by some analysts. Lo and MacKinlay (1988) demonstrated that the serial correlation of share prices is significantly significant. Therefore, there is a possibility of short-term returns on share prices when investors realize that share prices move consequently in the same direction. Studying the American market with high-frequency data for the S&P 500 index, Peters (1994) found a persistent time series with strong autocorrelation. Findings from other recent studies, discussed in the literature review, are also consistent with our present results.

What may be the reasons for the mixed empirical evidence for the efficient market hypothesis?¹¹

We do not have a solid answer but believe that the conflicting findings may be a result of discrepancies in the datasets used in prior studies. From our prior experience, estimation results using data from the same stock indices obtained from different databases can sometimes be quite dissimilar, perhaps due to various methodologies used in constructing the data series. Econometric methods employed in a given study can also play a role. Bhargava (2014) demonstrated that certain approaches in testing for random walk (such as those by Lo and Andrew’s variance ratio and related tests) can lead to erroneous results. Our study, we believe, mitigated some of these shortcomings by employing a comprehensive battery of highly regarded tests on an authoritative database. It is surprising that this was the first time, to our knowledge, data from the WRDS stock price indices were used to examine this issue.

6. Concluding Remarks

The main rationale for our research was that previous studies had found mixed results with regard to the efficient market hypothesis. We set out to explore this topic for the USA, the UK and Japan with a recent dataset and improved statistical methods. We contributed to the existing literature by employing a comprehensive battery of tests including several high-power multiple-break unit root and novel spectral tests. We further computed the abnormal returns using the break dates captured in the models. We then linked those abnormal profits to their associated economic events. We found that stock market indices in the USA, the UK and Japan are generally not efficient. While our results are in line with a number of recent studies, they do not support the findings of several earlier studies reviewed in the paper. Therefore, definitive empirical evidence for mean reversions in highly developed markets remains elusive. It will be interesting to extend the present study to include market indices in other advanced countries in future studies. Finally, based on the findings in this study, it may be concluded that investors could possibly be able to earn arbitrage profits due to market inefficiency even in highly developed stock markets.

Author Contributions: Conceptualization, J.N.; methodology, J.N. and W.-X.L.; software, J.N.; validation, J.N., W.-X.L. and C.C.-S.C.; formal analysis, J.N.; investigation, J.N. and W.-X.L.; resources, J.N.; data curation, J.N.; writing—original draft preparation, J.N.; writing—review and editing, J.N., W.-X.L. and C.C.-S.C.; visualization, J.N.; supervision, J.N.; project administration, J.N.; funding acquisition, J.N. and W.-X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is available upon request from the authors.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ These tests allow for more than one structural break in the data and, if not accounted for, can lead to misleading results (Lumsdaine and Papell 1997; Lee and Strazicich 2003; Narayan and Popp 2010; Ender and Lee 2012).
- ² We strongly suggest the readers refer to the original papers for detailed derivations of the models and test statistics. Due to space limitations and the large number of tests examined in this study, it is not practical to discuss each of them in detail.
- ³ There is another model based on the ergodic theorem stating that past and present probability distributions define the probability distribution, which will help forecast future market prices. The ergodic principle posits that the future is predetermined by the existing variables such as market fundamentals. Therefore, it is possible to forecast the future by analyzing the present and past data. If the system is nonergodic, on the other hand, the probability distributions of past and present do not provide a statistically reliable estimate for the probability of future events. A reviewer commented that stock prices appear to be random, yet they are “chaotic” in reality. This presents a challenge for the random walk model. Klinkova and Grabinski (2017) and Grabinski and Klinkova (2019) showed that using arithmetic means in chaotically varying quantities does not always yield results to random variations and that the “ultimate” financial model is not possible. Furthermore, ergodicity can be assumed in random variations but, generally, not in chaotic ones.
- ⁴ We selected high-impact and widely cited tests (most of which were originally published in elite journals in the fields of econometrics, statistics, finance and economics) to be used in our study to avoid the “kitchen-sink” approach.
- ⁵ To conserve space, we reported the results for two lags since the results were essentially the same for any of these methods.
- ⁶ Please refer to Wei (2018) and Ronderos (2014) for detailed discussions of the tests in this section.
- ⁷ To conserve space, we did not report the results from all the tests conducted in this study discussed in the Data and Methodologies section, especially when the vast majority of the findings were similar. Rather, we focused on the more interesting and important test results. In addition to the reported tests, we completed a variety of older random walk tests such as the Brock et al. (1996), various versions of variance ratio, runs and autocorrelation tests as in several of the reviewed articles and found the results were essentially unchanged (and did not report them in the Results section). The complete results are available from the authors upon request.

- 8 The correlation coefficients between the WRDS indices and those of Compustat are between 0.95 and 0.98 for the countries in our sample.
- 9 An anonymous reviewer noted that one typically wants to show that the measured results are stronger with a statistical significance when there is a null hypothesis or placebo. In many cases, the null hypothesis is also a result of observation. As such, it has a distribution. Including both distributions, consequently, changes the way one proves statistical significance. In a recent study, [Tormählen et al. \(2021\)](#) showed that in order to obtain identical significance, it may be necessary to perform twice as many experiments than in a setting where the placebo distribution is ignored. They also showed that statistical significance may be inaccurate due to “misuse” of the central limit theorem.
- 10 The specification with three and more structural breaks was tested. However, our statistical software only found two breaks. Furthermore, the results remained similar regardless of the number of lags employed.
- 11 We thank an anonymous referee for his/her many stimulating questions, including this one.

References

- Amihud, Yakov. 2002. Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Market* 5: 31–56. [\[CrossRef\]](#)
- Ang, Andrew, and Geert Bekaert. 2007. Return predictability: Is it there? *Review of Financial Studies* 20: 651–707. [\[CrossRef\]](#)
- Baker, Malcolm, and Jeffrey Wurgler. 2006. Investor sentiment and the cross-section of stock returns. *Journal of Finance* 61: 1645–80. [\[CrossRef\]](#)
- Baker, Malcolm, and Jeffrey Wurgler. 2007. Investor sentiment in the stock market. *Journal of Economic Perspectives* 21: 129–51. [\[CrossRef\]](#)
- Bekaert, Geert, and Robert James Hodrick. 1992. Characterizing predictable components in excess returns on equity and foreign exchange markets. *Journal of Finance* 47: 467–509. [\[CrossRef\]](#)
- Bekaert, Geert, Campbell Russell Harvey, and Christian Lundblad. 2007. Liquidity and expected returns: Lessons from emerging markets. *Review of Financial Studies* 20: 1783–831. [\[CrossRef\]](#)
- Bertone, Stephen, Imants Paeglis, and Rahul Ravi. 2015. (How) has the market come more efficient? *Journal of Banking & Finance* 54: 72–86.
- Bhargava, Alok. 1986. On the theory of testing for unit roots in observed time series. *Review of Economic Studies* 53: 369–84. [\[CrossRef\]](#)
- Bhargava, Alok. 2014. Firms’ fundamentals, macroeconomic variables and quarterly stock prices in the US. *Journal of Econometrics* 183: 241–50. [\[CrossRef\]](#)
- Borges, Maria Rosa. 2010. Efficient market hypothesis in European stock markets. *The European Journal of Finance* 16: 711–26. [\[CrossRef\]](#)
- Brock, William, W. Davis Dechert, Blake Lebaron, and Jose Scheinkman. 1996. A test for independence based on a correlation dimension. *Econometric Review* 15: 197–235. [\[CrossRef\]](#)
- Busse, Jeffrey A., Amit Goyal, and Sunil Wahal. 2010. Performance and persistence in institutional investment management. *Journal of Finance* 65: 765–90. [\[CrossRef\]](#)
- Campbell, John Young, and Yasushi Hamao. 1992. Predictable stock returns in the United States and Japan: A study of long-term capital market integration. *Journal of Finance* 47: 43–69. [\[CrossRef\]](#)
- Campbell, John Young, and Robert James Shiller. 1988. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1: 195–228. [\[CrossRef\]](#)
- Carhart, Mark. 1997. On persistence in mutual fund performance. *Journal of Finance* 52: 57–82. [\[CrossRef\]](#)
- Chen, Nai-Fu, Richard Roll, and Stephen Ross. 1986. Economic forces and the stock market. *Journal of Business* 59: 383–403. [\[CrossRef\]](#)
- Chow, K. Victor, and Karen C. Denning. 1993. A simple multiple variance ratio test. *Journal of Econometrics* 58: 385–401. [\[CrossRef\]](#)
- Collins, Robert M. 1996. The Economic Crisis of 1968 and the Waning of the “American Century”. *The American Historical Review* 101: 396–422. [\[CrossRef\]](#)
- De Long, James Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert James Waldmann. 1990. Positive feedback investment strategies and destabilizing rational speculation. *Journal of Finance* 45: 379–95. [\[CrossRef\]](#)
- Dickey, David Alan, and Wayne Arthur Fuller. 1979. Distribution of the estimates for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74: 427–31.
- Dickey, David Alan, and Wayne Arthur Fuller. 1981. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49: 1057–72. [\[CrossRef\]](#)
- Durusu-Ciftci, Dilek, Mustafa Serdar Ispir, and Hakan Yetkiner. 2017. Financial development and economic growth: Some theory and more evidence. *Journal of Policy Modeling* 39: 290–306. [\[CrossRef\]](#)
- Elliott, Graham, Thomas J. Rothenberg, and James Harold Stock. 1996. Efficient tests for an autoregressive unit root. *Econometrica* 64: 831–36. [\[CrossRef\]](#)
- Ender, Walter, and Junsoo Lee. 2012. A unit root test using a fourier series to approximate smooth breaks. *Oxford Bulletin of Economics and Statistics* 74: 574–99. [\[CrossRef\]](#)
- Fama, Eugene Francis, and G. William Schwert. 1977. Asset returns and inflation. *Journal of Financial Economics* 5: 115–46. [\[CrossRef\]](#)
- Fama, Eugene Francis, and Kenneth Ronald French. 1988. Permanent and temporary components of stock prices. *Journal of Political Economy* 96: 246–73. [\[CrossRef\]](#)
- Fama, Eugene Francis, and Kenneth Ronald French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47: 427–65. [\[CrossRef\]](#)

- Fama, Eugene Francis, and Kenneth Ronald French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56. [\[CrossRef\]](#)
- Fama, Eugene Francis, and Kenneth Ronald French. 2010. Luck versus skill in the cross section of mutual fund returns. *Journal of Finance* 65: 1915–47. [\[CrossRef\]](#)
- French, Kenneth Ronald, and Richard Roll. 1986. Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics* 17: 5–26. [\[CrossRef\]](#)
- Goetzmann, William, Roger Ibbotson, and Liang Peng. 2001. A new historical database for the NYSE 1815 to 1925: Performance and predictability. *Journal of Financial Markets* 4: 1–32. [\[CrossRef\]](#)
- Golez, Benjamin, and Peter Koudijs. 2018. Four centuries of return predictability. *Journal of Financial Economics* 127: 248–63. [\[CrossRef\]](#)
- Grabinski, Michael, and Galiya Klinkova. 2019. Wrong use of average implies wrong results from many heuristic models. *Applied Mathematics* 10: 605–18. [\[CrossRef\]](#)
- Grossman, Sanford, and Joseph Stiglitz. 1980. On the impossibility of informationally efficient markets. *American Economic Review* 70: 393–408.
- Horst, Ulrich, and Jan Wenzelburger. 2008. On non-ergodic asset prices. *Economic Theory* 34: 207–34. [\[CrossRef\]](#)
- Johansen, Anders, and Didier Sornette. 2020. Condensed Matter and Complex Systems. *European Physical Journal B* 17: 319–28. [\[CrossRef\]](#)
- Klinkova, Galiya, and Michael Grabinski. 2017. Conservation laws derived from systemic approach and symmetry. *International Journal of Latest Trends in Finance and Economics* 7: 1307–12.
- Lee, Junsoo, and Mark C. Strazicich. 2003. Minimum lagrange multiplier unit root test with two structural breaks. *Review of Economics and Statistics* 85: 1082–89. [\[CrossRef\]](#)
- Lettau, Martin, and Stijn Van Nieuwerburgh. 2008. Reconciling the return predictability evidence. *Review of Financial Studies* 21: 1607–52. [\[CrossRef\]](#)
- Lo, Andrew Wen-Chuan, and A. Craig MacKinlay. 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies* 1: 41–66. [\[CrossRef\]](#)
- Lumsdaine, Robin L., and David H. Papell. 1997. Multiple trend breaks and the unit-root hypothesis. *Review of Economics and Statistics* 79: 212–18. [\[CrossRef\]](#)
- Mitra, Subrata, Manojit Chattopadhyay, Charan Parikshit, and Bawa Jaslene. 2017. Identifying periods of market inefficiency for return predictability. *Applied Economics Letters* 24: 668–71. [\[CrossRef\]](#)
- Narayan, Paresh, and Stephan Popp. 2010. A new unit root test with two structural breaks in level and slope at unknown time. *Journal of Applied Statistics* 37: 1425–38. [\[CrossRef\]](#)
- Narayan, Paresh, and Russell Smyth. 2007. Are shocks to energy consumption permanent or temporary? Evidence from 182 countries. *Energy Policy* 35: 333–41. [\[CrossRef\]](#)
- Ng, Serena, and Pierre Perron. 2001. Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69: 1519–54. [\[CrossRef\]](#)
- Nyberg, Henri, and Harri Pönkä. 2016. International sign predictability of stock returns: The role of the United States. *Economic Modelling* 58: 323–38. [\[CrossRef\]](#)
- Perron, Pierre. 1989. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57: 1361–401. [\[CrossRef\]](#)
- Peters, Edgar E. 1994. *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*. New York: John Wiley & Sons Inc.
- Phillips, Peter C. B., and Pierre Perron. 1988. Testing for a unit root in time series regression. *Biometrika* 75: 335–46. [\[CrossRef\]](#)
- Rapach, David E., Jack K. Strauss, and Guofu Zhou. 2013. International stock return predictability: What is the role of the United States? *Journal of Finance* 68: 1633–62. [\[CrossRef\]](#)
- Ronderos, Nicolas. 2014. Spectral Analysis Using Eviews. Available online: <https://www.eviews.com/Addins/SpectralAnalysis.aipz> (accessed on 18 January 2022).
- Schwert, G. William. 1989. Tests for unit roots: A monte carlo investigation. *Journal of Business and Economic Statistics* 7: 147–60.
- Shiller, Robert James. 1981. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review* 71: 421–36.
- Shiller, Robert James. 1982. Consumption, asset markets, and macroeconomic fluctuations. *Carnegie-Rochester Conference Series on Public Policy* 17: 203–38. [\[CrossRef\]](#)
- Solnik, Bruno. 1993. The performance of international asset allocation strategies using conditioning information. *Journal of Empirical Finance* 1: 33–55. [\[CrossRef\]](#)
- Tormählen, Maike, Galiya Klinkova, and Michael Grabinski. 2021. Statistical Significance Revisited. *Mathematics* 9: 958. [\[CrossRef\]](#)
- Urquhart, Andrew, and Robert Hudson. 2013. Efficient or adaptive markets? Evidence from major stock markets using very long run historic data. *International Review of Financial Analysis* 28: 130–42. [\[CrossRef\]](#)
- Urquhart, Andrew, and Frank McGroarty. 2016. Are stock markets really efficient? Evidence of the adaptive market hypothesis. *International Review of Financial Analysis* 47: 39–49. [\[CrossRef\]](#)
- Verheyden, Tim, Lieven De Moor, and Filip Van den Bossche. 2015. Towards a new framework on efficient markets. *Research in International Business and Finance* 34: 294–308. [\[CrossRef\]](#)

Wei, William W. S. 2018. *Time Series Analysis*, 2nd ed. Boston: Addison Wesley.

Zivot, Eric, and Donald Wilfrid Kao Andrews. 1992. Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics* 10: 251–70.