EMERGING MARKETS FINANCE AND TRADE
https://doi.org/10.1080/1540496X.2021.1949981

A Study of Market Efficiency in Emerging Markets Using Improved Statistical Techniques

James Nguyen and Richard Parsons

Economics and Finance, Texas A&M University Texarkana, Texarkana, Texas, USA

ABSTRACT
This article helps resolve the current unsatisfying and inclusive studies covering the efficiency of stock markets in developing countries. Previous studies have used limited data and partial statistical tests. We use a large, unique data set, across 12 countries, and a comprehensive set of traditional and recent statistical methods as well as powerful multiple-break unit root and spectral analysis tests, many of which have never been used to evaluate the efficient market hypothesis (EMH) in emerging markets. Our results confirm the rejection of the EMH for emerging markets. Our findings have important implications for investors and policy makers, suggesting the possibility for excess profits in these markets.

KEYWORDS
Developing markets; efficient market hypothesis; BDS test; multi-break unit root tests; spectral analysis

JEL
C32; G14; G15; O16

1. Introduction

The idea of efficient markets has been a central topic in financial theory since Eugene Fama introduced it in 1970. The efficient market hypothesis (EMH) states that financial asset prices entirely reflect all available information, making it impossible for investors to beat the market. The theory has received significant support from academics: Shiller (1981), Carhart (1997), Lettau and Van Nieuwerburgh (2008), Fama and French (2010), Busse, Goyal, and Wahal (2010), and Bertone, Paeglis and Ravi (2015). The award of the Nobel Prize in Economics to Richard Thaler in 2017 has helped reignite this debate. Thaler, one of the founders of "behavioral finance" has put the notion of 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, Moor, and Bossche 2015).

After extensive research of the EMH in developed markets, a widespread but not complete consensus exists that developed markets tend toward efficiency although there are periods of informational inefficiency and periods of speculative bubbles (behavioral finance), (French and Roll (1986), De Long and Becht (1992), Carhart (1997), Fama and French (2010), Busse, Goyal, and Wahal (2010), Bertone, Paeglis, and Ravi (2015).

When it comes to EMH, there has been much less study of undeveloped markets. Developing markets are thought to have less efficient information flow and less accurate institutional information and therefore more friction and asymmetric information. Bekaert and Harvey (2002) demonstrate evidence on relative inefficiency of emerging markets by showing that liquidity plays a major role. Griffin, Kelly, and Nardari (2010) show that in developing markets commonly used efficiency tests do not control for the information environment. Kaminsky and Schmukler (2002) show that developing markets are less stable in the long run. Moreover, Stiglitz (2014) documents that developing markets are very sensitive to new information, both facts and perception with little underlying relationship to

CONTACT Richard Parsons rparsons@tamut.edu Economics and Finance, Texas A&M University Texarkana, 7101 University Ave, Texarkana, Texas 75501, USA

fundamentals. Given these informational shortcomings of developing markets, the constant and unsteady flow of new information presents significant investor challenges.

Previous studies of EMH in developing countries rely on one or two countries and limited statistics. Our article contributes to the literature by testing a relatively large number of developing markets rather than one or two isolated countries and by employing a more complete and powerful battery of statistical techniques. In this study, time series data over a number of years (for some countries over 20) instead of one or a few isolated countries based on geographical regions are used. The 12 countries[1] in our data set are classified as "developing markets" by two authoritative sources. Our study also complements those that examine this issue for major stock markets, especially the study of US, UK, and Japanese stock markets by Urquhart and McGroarty (2016), Urquhart and Hudson (2013) and that of Borges (2010) and others. We utilize a set of well-known, highly regarded statistical tests to explore this topic, in addition to the traditional methods (PP, Runs Test, ADF/Augmented ADF, variance ratio tests, and autocorrelation tests). Specifically, we employ in this article important tests such as those by Elliott, Rothenberg, and Stock (1996), Ng and Perron (2001), various versions of variance ratio tests, Brock, Dechert, and Scheinkmanal (1996), *multiple*-break unit root tests by Lumsdaine and Papell (1997), Lee and Strazicich (2003)[2] and Narayan and Popp (2010) – some of which have not been widely used in this line of research. Further, we adopt recent techniques from spectral analysis commonly used in Electrical Engineering and Physics to obtain additional insights and to increase the robustness of our results. To our knowledge, this is the first time these tests are used to examine stock market efficiency in the 12 emerging markets in our study, despite their numerous advantages. Finally, our study is the first to employ a unique, authoritative dataset provided by WRDS, described in detail in Section 2.

By way of preview, we find that the 12 emerging stock market indices in our study generally exhibit mean reversions. The single-break unit root test results are not as definitive as those derived from other methods and that spectral tests confirm most of the results from other tests. The general finding of market inefficiency may indicate more pronounced information asymmetry, limited competition, and underdeveloped financial and banking systems within these countries. The article is organized as follows. Section 2 reviews the relevant, recent studies. Section 3 discusses the data and methodology. Section 4 presents the empirical results. Section 5 concludes the study with some remarks.

2. A Brief Literature Review

Recently, there have been several empirical studies which challenged the original or strong form of EMH and proposed adjustments to the theory, including the "weak form" and "adaptive EMH" Lo (2004). A number of empirical tests have documented that return predictability varies over time showing periods of inefficient markets and periods of efficient markets, for example, Mitra et al. (2017), Lim and Brooks (2006), Smith (2011), Lim, Luo, and Kim (2013), and Urquhart and Hudson (2013).

The EMH relies upon the idea of symmetric and complete information so that investors can make informed and rational decisions. Stiglitz (2014) reminds us that financial markets are inherently about information. However, Grossman and Stiglitz (1976) point out that symmetric and complete information is a troubling assumption and goes as far as saying that "because information is costly, prices cannot perfectly reflect the information which is available, since if it did, those who spent resources to obtain it would receive no compensation" (p. 405). Moreover, Shiller (1981, 2) tells us, "Markets in which there is imperfect and asymmetric information are not in general constrained Pareto efficient." These information discussions become critical as we move on to the study of developing markets.

A limited number of studies have recently tested EMH in one or two developing markets at a time. However, these studies have drawbacks of finite samples and are inconclusive. Hiremath and Kumari (2014) supported the idea of adaptive markets in India showing that the market switched between periods of efficiency and inefficiency but overall was moving toward efficiency. Hadhri and Ftiti (2017) found that developing market asset returns could be robustly modeled in the Middle East, North Africa, and Latin America, but could find no standard model of return. Kheradyar, Ibrahim and Mat

Nor (2011) showed that ratio analysis was a valid predictor in the Malaysian market. In addition, Arltova (2000) who used variance ratio tests to show the changing volatility over time rejected the random walk for the Czech market. The studies of developing market efficiency have been limited in depth and breadth and are not conclusive. Developing equity markets have sprung up all around the world over the last several decades and are now a critical part of diversified investing. Borges (2011) studied the Spanish stock market and found mixed evidence of market efficiency. Seth and Sharma (2015) examined the informational efficiency and integration for select Asian and US markets and found that the markets in question are weakly inefficient. Malafeyev, Awasthi, and Kambekar (2017) studied the issue in the Chinese and Indian equity markets and found that these markets do not exhibit (weak-form) market efficiency. More recently, Xiong et al. (2019) showed that the Chinese stock market is efficient but is better described by the Adaptive Market Hypothesis which attempts to reconcile the theory of market efficiency and behavioral finance.

The approach for testing market efficiency has always been centered around the random walk. A random walk implies that the movement of a stock from its beginning point is a stochastic process and independent of past movements and therefore unpredictable. In probability theory, a martingale sequence is often used to describe a generalized random walk. Past studies have primarily used autoregressive time series tests to determine the predictability of a trend. In this paper, we will use six classes of sophisticated test as outlined below[3]. These are discussed in detail in Section 3 of this article. These six tests are One, autoregressive time series test (test of randomness including the runs test). Two, variance-ratio methodology which has been popular in recent years (Charles and Darné, 2009). Three, unit root tests (Dickey–Fuller, Phillip Perron, Ng–Perron, Elliot–Rothernbert–Stock). Four, Brock Dechert and Scheinkman (1996) test of independence. Five, multiple break Unit Root Test: Lumsdaine and Papell (1997), Lee and Strazicich (2003) and Narayan and Popp (2010). Six, spectral analysis (first time this has been used to study EMH in developing markets).

3. Data and Methodology

The efficient market hypothesis (weak-form) states that stock prices move in a random walk fashion or past prices cannot be used to predict future prices. To evaluate this hypothesis, returns are examined using tests for independence. While we use a variety of methods as traditionally employed in prior studies (autocorrelation/partial correlations, variance ratios, BDS, runs, ADF, Zivot–Andrews, and various versions of simple unit root tests), we focus our attention to the recent and more powerful techniques (and present the results associated with these tests). This is because Monte Carlo studies from Elliot, Rothenberg, Stock (ERS), and NG–Perron, among others, have found that traditional tests perform extremely poorly. Examining a sample correlogram to check for unit roots, for instance, is very inaccurate as what appears as a unit root to one person may look like a stationary process to another person (Enders 2008). Consequently, we focus on the results from the more powerful tests presented in the tables. Particularly, to preserve journal space, we primarily discuss the results from the variance ratio, ADF, ERS-GLS, Ng–Perron, BDS, simple/multiple unit root, and spectral tests. Below is a brief review of these tests. Readers are encouraged to refer to the original articles for a detailed understanding of these tests (cited in the reference section). Before performing the aforementioned tests, we also examine every series for autocorrelations via its associated correlograms for the presence of autocorrelation and perform the usual diagnostic tests as typically done in the literature. We then proceed to perform our main battery of tests – as discussed. If the traditional test results are similar to those found in the main tests, we do not report them. If there are conflicting results, we discuss them in the corresponding country analyses. We do not report the full results for each of the tests (they are available upon request from the authors) to conserve journal space and to simplify the exposition.

Our data set is obtained from the Wharton Research Data Services (WRDS) country price index database, globally known for providing reliable data for research[4]. A major advantage of using this database is that all the price series have a consistent data format. Our sample contains daily data for 12 emerging countries which are Brazil, Chile, Columbia, Hong Kong, India, Indonesia, Korea, Malaysia, Mexico, Philippines, Taiwan, and Thailand. Compustat Global – Security Daily was used to construct the indices, which are market-capitalization weighted. The portfolio is rebalanced annually at the end of the last trading day of June, for each country. Observations are removed if the market capitalization is not positive or if the exchange information is missing. For firms that have multiple issues, the issue with the largest market capitalization is chosen. Also, a security has to be in the top 50% of the market capitalization of that country and have to trade in the stock exchanges located within the country in question. The currency of the security price has to be consistent with its ISO currency code. Lastly, only common ordinary shares are included in the indice[5].

There is no formal list of emerging markets. Varieties of lists have appeared in various publications. We examined nine different contemporary lists but found that they were inconsistent, probably due to the fact that very different sets of criteria were used in constructing the lists. Therefore, our approach is to use the combined list from two very rigorous studies. Kaminsky and Schmukler's (2002) World Bank article lists 11 countries as emerging stock markets and Henry's (2000) Journal of Finance article lists the same 11 countries but also includes India as an addition. Consequently, we decided to use the data for these 12 countries for this study.

4. 3.a *Variance Ratio Tests*

These tests, originally due to Lo and MacKinlay (1988), have been shown to be more powerful and reliable than ADF tests and are robust to heteroscedasticity. The Lo-MacKinlay test 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 choose 2,4,8,16 periods as typically done in the literature. The variance ratio of Lo and MacKinlay tests whether the variance ratio is equal to 1 for a particular holding period. For each country, we present its variance ratio, its Chow and Denning (1993) joint maximum {z} statistic (since we choose more than 1 period) and its associated *p*-values (we do 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 5% significance level). Note that, for a given set of test statistics, the random walk hypothesis is rejected if any one of the variance ratios is considerably dissimilar than one. As an additional robustness check, we also perform an improved, more powerful version of the standard variance ratio test, the Wright (2000) variance ratio test, which is obtained by computing the Lo and MacKinlay homoscedastic test statistic using ranks and rank scores instead of the original data. The results of this test are not reported, to conserve space, if they are similar to those obtained from the Lo and MacKinlay tests. Table 2 contains our VR test results.

5. 3.b *Unit Root Tests*

We present the main results from the ADF, ERS DF-GLS, and Ng–Perron tests of unit root in Table 3. For ease of interpretation, we report results with the stated significance levels.

ADF test (1981): this is our baseline test which aims to examine if a series is stationary or random walk (unit root), mainly for comparison purposes. The null hypothesis of unit root is rejected if the test statistics (presented in Table 3) are less than their associated critical values. The problem with ADF method is the selection of lag length (Schwert 1989). We, therefore, use (automatic) Akaike's Information Criterion to select the optimal lag length (to ensure that the residual is white noise) in order to mitigate this issue. We also perform the Phillips and Perron (1988) test, a more powerful test than the ADF test, but with more size distortions, and do not present the results (as they almost always yield the same answers as its ADF counterparts).

ERS test (1996): this is basically a modified ADF test where Elliot, Rothenberg, and Stock (ERS) show that their DF-GLS test (shown in Table 3) has the power function close to the point optimal test which has better power properties. This test not only provides higher power than ADF and PP tests but

Table 1. Data.

	Number				Jarque–Bera	Dates	
	Observations	Mean	Kurtosis	Skewness	Normality	Begin	End
Taiwan	6684	0.000143	5.727469	−0.119433	No	7/4/1988	12/9/2015
Thailand	6302	0.000315	14.30094	0.552215	No	7/4/1988	12/10/2015
Brazil	4672	0.000496	18.15089	−0.260931	No	7/3/1995	12/9/2015
Philippines	4476	0.000353	12.23289	0.199781	No	7/3/1995	12/9/2015
Chile	2892	0.000508	19.23079	0.453684	No	3/11/2002	12/9/2015
Korea	2815	0.000377	8.114345	−0.258634	No	3/11/2002	12/9/2015
Columbia	2102	0.000475	9.096227	−0.114374	No	7/1/2005	12/10/2015
Malaysia	2038	0.000327	16.50680	−1.450418	No	7/1/2005	12/11/2015
Indonesia	1964	0.000275	14.18709	0.038857	No	7/2/2005	12/12/2015
Hong Kong	1923	−0.000006	8.030026	−0.278254	No	7/3/2005	12/13/2015
Mexico	5313	0.000783	9.085962	0.255951	No	7/4/1988	12/10/2015
India	1923	0.000286	14.22689	0.059857	No	7/5/2005	12/9/2015

Table 2. Variance ratio tests.

	Period				
VR Test	2	3	8	16	Max (Z)
Taiwan	0.5111*	0.2741*	0.1272*	0.0666*	22.5990*
Thailand	0.5681*	0.3026*	0.1494*	0.0751*	12.5740*
Brazil	0.5435*	0.2755*	0.1321*	0.0694*	11.5070*
Philippines	0.5875*	0.2984*	0.1416*	0.0741*	13.0414*
Chile	0.6148*	0.3050*	0.1444*	0.0809*	8.3075*
Korea	0.5153*	0.2558*	0.1286*	0.0625*	13.2069*
Columbia	0.5226*	0.2701*	0.1281*	0.0630*	8.8034*
Malaysia	0.5630*	0.2790*	0.1361*	0.0665*	7.1475*
Indonesia	0.5182*	0.2393*	0.1395*	0.0674*	9.8123*
Hong Kong	0.5042*	0.2475*	0.1237*	0.0642*	11.0360*
Mexico	0.5778*	0.3069*	0.1610*	0.0841*	30.3062*
India	0.5007*	0.2699*	0.1443*	0.0698*	6.9784*

>1% significance.

Table 3. Unit root tests.

			NG–Perron			
	Augmented Dickey–Fuller	ERS DF-GLS	Mza	MZt	MSB	MPT
Taiwan	−77.6224*	−3.8426	−10.5718**	−2.2810*	0.2518**	0.2391**
Thailand	−12.0243*	−0.1982*	−440.5410	−14.8400	0.0337*	0.0570*
Brazil	−63.3250*	0.0159	−10.5718**	−2.2810**	0.2158**	2.3909***
Philippines	−14.0095*	−10.0532*	−104.0590*	−7.2128*	0.0693*	0.2361*
Chile	−44.8138*	−13.8356*	−4.3400	−1.4737	0.3392	5.6399
Korea	−52.1715*	−3.7373*	−6.7273***	−1.8242***	0.2712	3.6777***
Columbia	− 7.7248*	−0.9744	−2.0573	−0.9664	0.4697*	11.4565
Malaysia	−9.0542*	−2.4253**	−3.8887	−1.3027	0.3350	6.3888
Indonesia	−8.3226*	−6.8004*	−32.3919*	−3.9809*	0.1229*	0.8907*
Hong Kong	−7.9744*	−5.1664*	−13.0899**	−2.5581**	0.1954**	1.8726**
Mexico	−65.2561*	0.0971*	−2539.9900*	−35.6369*	0.0140*	0.0097*
India	−8.0881*	−8.0764*	−55.9904*	−5.2905*	0.0945*	0.4389*

*>1% significance.
**>5% significance.
***>10% significance.

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 is the first time the ERS tests are used to examine the market efficiency hypothesis, at least for our sample of countries.

Ng–Perron test (2001): using the procedure in ERS to create efficient versions of the modified PP tests of Perron and Ng (1996), Ng and Perron (2001) show that these tests do not have the same serious size distortions of the PP tests (used in many studies reviewed in the article) for errors with large autoregressive and moving average roots. As a result, they can give much higher power than PP tests. Ng and Perron construct four test statistics (shown in Table 3) which are based on the PP tests (MZa and MZt statistics), the Bhargava (1986) (MSB) statistic and the ERS Point Optimal statistic (MPT). We use the modified AIC for lag selection, as suggested by the authors, to maximize power. Interpretations of the results for these tests are similar to those of the ADF tests discussed above. To our knowledge, this is the first time these tests are used to examine the market efficiency hypothesis for our sample of countries.

6. 3.c *BDS Test (1996)*

This is perhaps the most powerful (nonlinear) tests for detecting serial dependence in time series data, due to Brock et al. (1996). This is essentially a nonlinear analog of the Box Pierce Q statistic used in ARIMA analysis. It is important to include this test as other linear tests may fail to detect nonlinear predictability (Amini, Hudson, and Keasey 2010). De Lima (1998) was able to find shifts in the distribution of stock market data when autoregressive conditional heteroscedastic filters were unable to do so. A number of studies have found evidence of the movement of asset returns. It 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 present results using different dimensions (m = 2 to 6) and e = 0.5 (shown in Table 4). The distance e is 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. Also, we let e vary from 0.50 to 2 as suggested in the econometric literature (the higher this value, the lower the power of the test). The results for tests where e is 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 choose the option of calculating the bootstrapped p-values for the test statistic using various repetitions to increase the accuracy of the p-values (results are not shown in Table 3 as they are qualitatively similar to those from the standard tests). It should be noted that, prior to performing each BDS test, we remove any linear dependence from the data by fitting a GARCH (p,q) model and apply the BDS test to the corresponding residuals. As an additional test, we also apply the BDS test to the original data and discuss the results if there are differences.

Table 4. BDS tests.

	BDS by Dimension				
	2	3	4	5	6
Taiwan	0.0234*	0.0514*	0.0776*	0.0855*	0.0928*
Thailand	0.0315	0.0610	0.0820	0.0943	0.1001*
Brazil	0.0174*	0.0361*	0.0503*	0.0579*	0.0612*
Philippines	0.0221*	0.0440*	0.0600*	0.0699*	0.0741*
Chile	0.0240*	0.0468*	0.0615*	0.0691*	0.0721*
Korea	0.0127*	0.0338*	0.0532*	0.0671*	0.0753*
Columbia	0.0258*	0.0448*	0.0573*	0.0623*	0.0621*
Malaysia	0.0276*	0.0566*	0.0764*	0.0882*	0.0928*
Indonesia	0.0238*	0.0500*	0.0678*	0.0795*	0.0850*
Hong Kong	0.0174*	0.0380*	0.0546*	0.0650*	0.0692*
Mexico	0.0226*	0.0456*	0.0634*	0.0750*	0.0817*
India	0.0236*	0.0489*	0.0704*	0.0876*	0.0963*

*>1% significance.
**>5% significance.
***>10% significance.

6.1. 3.d Multiple-Break Unit Root Tests

One of the novel contributions of our study is the inclusion of multiple-break unit root tests by Lumsdaine and Papell (1997) and Lee and Strazicich (2003) and Narayan and Popp (2010). To our knowledge, these tests were not used in any of the previous studies examining stock return predictability in emerging markets, particularly our sample of countries. It is a good idea to allow for multiple breaks as many emerging countries such as those in our data set have experienced many economic reforms in the last several decades during their transitions to market economies. The main limitation of the Zivot–Andrews (1992) unit root test is that it allows only for one break in the data[6]. The LP multiple unit root tests allow for more than one (unknown) break point in either the trend, the intercept or both the trend and intercept of the data. We use 2,4 and 6 lags for the base model as well as automatic lag selections using the AIC and BIC[7]. The LS unit root test is another alternative multiple-break test. We employ both the *Crash* Model to allow for a sudden change in level but no change in the trend and *Break* Model to account for simultaneous changes in the level and trend. The null hypothesis for both LP and LS tests is that there is a unit root in the data. Thus, if the null hypothesis is rejected (denoted as "Y" for "Yes" – to facilitate the exposition in Table 5), the return series is predictable and vice versa (denoted as "N" for "No"). It is important to note that the critical values for LS test cannot be computed when the number of breaks is greater than 2 in practice. Lee and Strazicich (2003) show that their model outperforms that of Lumsdaine and Papell (1997) in simulations and that, unlike the LP unit root test, rejection of the null unambiguously implies trend stationary or return predictability in our case. They also show that the power of the tests increases substantially when two or more breaks are taken into account.

In 2013, Narayan and Popp published an article demonstrating that their 2010 test is significantly superior to the LP and LS tests in both size and power. Narayan and Popp (2010) studied an ADF-type innovational outlier unit root model that allows for two breaks in the level of a series and another that accounts for two breaks in the level and intercept. Breaks in their models are allowed under the null hypothesis as well as the alternative hypothesis. Using macroeconomic data for the United States, the authors show that their unit test has stable power, correct size and can accurately identify structural breaks[8].

It should be noted that these are computationally intensive methods when two or more breaks are selected, especially if the data set is fairly large (more than 500). For 500 observations, for example, it takes approximately 500 times as long to compute two breaks as one.

6.2. Spectral Analysis[9]

We perform three of the popular spectral analysis tests as suggested in textbooks by Warner (1998) and Wei (2018). We first examine the periodogram (a graph of the frequency in the horizontal axis and its amplitude in the ordinates (vertical axis)) – a tool used to help identify the dominant (important) periods, cyclical properties, or periodicities across different frequencies (high and low) in a series. The idea is that we divide the series into many parts, throw away unimportant parts of the data and stitch together the remaining parts. We look for peaks or hidden periodic components in the data. If the series seems very smooth, for example, then the values of the periodogram for low frequencies will be large relative to its other values (that is, the series has an excess of low frequencies and vice versa). For a random walk series, all the 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 non-stationary if the spectra are more prominent near frequency zero)[10]. To avoid spurious patterns, we employ the 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, it tests 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 the normalized integrated spectrum, we test the hypothesis if the observations from each of the series follow a white noise process.

Table 5. Multiple-break UNIT ROOT TESTs.

	Lumsdaine and Papell									Lee and Strazicich						Narayan and Popp					
	Breaks in Trend			Breaks in Intercept			Breaks in Trend and Intercept			Break Model			Crash Model			Breaks in Level			Breaks in Level and Trend		
Significance	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Taiwan	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Thailand	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Brazil	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Philippines	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Chile	N	N	N	N	N	Y	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Korea	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Columbia	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Malaysia	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Indonesia	N	N	N	N	N	Y	N	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Hong Kong	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Mexico	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
India	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

7. Empirical Results

Table 1 presents the summary statistics of the data. As found in many prior studies, all the series in the sample are not normally distributed on their associated J-B statistics. The stock markets in Taiwan, Thailand, and Mexico developed earlier, with starting dates in July 1998. Columbia, Malaysia, Indonesia, and Hong Kong began much later, beginning July 2005. The return series are quite similar, with the exception of Hong Kong, during the study period. We also examine the correlation matrix (results not shown) and observe that these return series are positively (and statistically significant) related, similar to those found in developed markets in several prior studies. The non-normality is significant because the very high kurtosis numbers indicate the likelihood of extreme returns. Also, the skewness numbers help reinforce that volatility with positive values indicates some extreme gains and many small losses, and vice versa for negative values.

Table 2 displays the variance ratio test results. First, as revealed in the table, the null hypothesis of random walk is decidedly rejected in all countries in the sample at the 1% significance levels for periods 2 through 16. Second, the Wright (2000) tests, not shown (as discussed), yield nearly identical results for all countries. These results are similar to several of those reported in some of the reviewed articles, using older data sets.

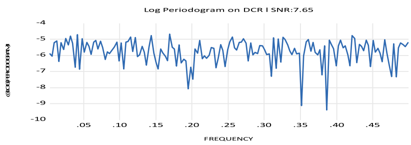
Table 3 shows the results for the unit root tests. The ADF tests unanimously reject the random walk hypothesis. However, the ERS DF-GLS and Ng–Perron tests provide less definitive results. Specifically, according to the ERS DF-GLS tests, Taiwan is the only country in the group in which the null hypothesis of unit root is not rejected. Taiwan is one of the countries in the sample with more advanced stock markets and highest GDP per capita. Based on the Ng–Perron tests, Malaysia and Chile are the only two countries where the null hypothesis is not rejected by all four associated tests. Overall, the Ng–Perron support the findings from the ADF and ERS DF-GLS tests with two exceptions. First, these tests are inconclusive for Thailand where two of the tests reject the null hypothesis and the remaining two tests fail to do so. Second, in the case of Columbia three of the Ng–Perron tests fail to reject the null hypothesis of nonstationarity. For this reason, we also examined the autocorrelation patterns of Columbia and notice that the hypothesis of "no autocorrelation" could not be rejected for the first eight lags. We also performed a panel unit root test (Levin, Lin, and Chu 2002) on these series and rejected the null of unit root.

Table 4 presents the results from the BDS tests. A quick glance at the table suggests that the null hypothesis of random walk (more precisely, the data follows an IID process) is rejected for all countries except Thailand where m = 6. These results are significant at the 1% level. As an experiment, we perform other traditional tests (PP, runs test, ADF, variance ratio tests, autocorrelation tests) and find that all of these tests reject the notion that this country's stock return series does not exhibit any predictability. We additionally apply the BDS test to the original data and obtain essentially the same results (not shown). Not surprisingly, we find that the results can sometimes be inconsistent. For instance, when examining the runs tests, we notice that the results are mixed – rejecting the RWH for about half of the countries. However, the main problem with this test is that it can detect serial correlation at lag one only. Overall, our results confirm that the isolated and limited findings from previous studies rejecting EMH for emerging markets can be more broadly accepted and more strongly supported.

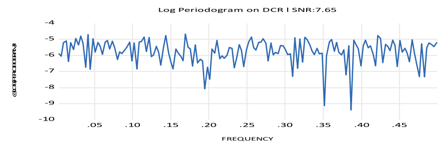
Table 5 displays the results of the LP, LS and NP multiple unit tests with two lags and two breaks as typically suggested in the econometric literature and as described in the corresponding section on page 12[11]. First, and most importantly, the null hypothesis of unit root (with two or more breaks) is rejected by all three tests at all significance levels for all countries except Chile and Indonesia. For Chile and Indonesia, the LP test results are mixed. The LP test results do reject the null at the 10% significance level when allowing for breaks in the intercept[12]. It is interesting to note that the NP unit root test results (which are claimed to be the most powerful) unanimously reject the null hypothesis at 1% significance level for all countries examined. Overall, these results are in accordance with those from many of the aforementioned tests.Table 6 summarizes the results of our spectral tests. Panels 1 through 12 display the Log Periodograms of the country series in the study[13]. First, as can be from the panels shown, most of strongest peaks occur at higher frequencies (except for South Korea) and the sinusoids are of varying strengths – as typically suggested by

Table 6. Spectral analysis.

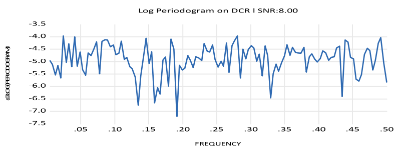
Panel 1 (Periodogram): Hongkong



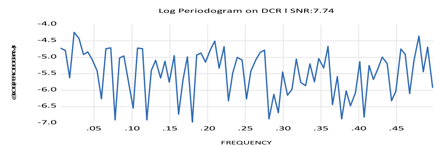
Panel 2 (Periodogram): Colombia



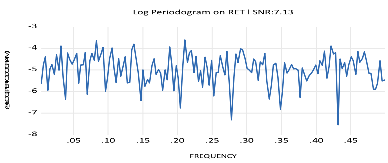
Panel 3 (Periodogram): Brazil



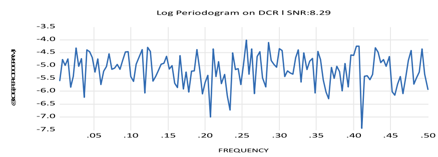
Panel 4 (Periodogram): Chili



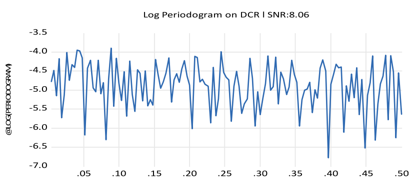
Panel 5 (Periodogram): Taiwan



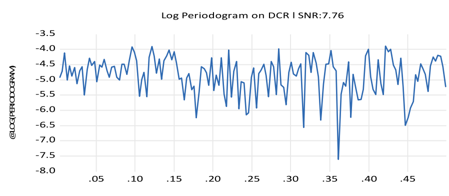
Panel 6 (Periodogram): Mexico



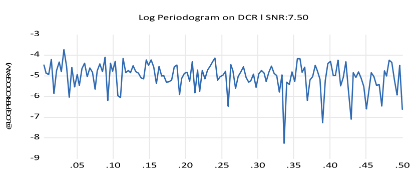
Panel 7 *(Periodogram):* India



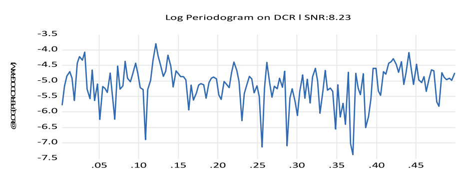
Panel 8 (Periodogram): Thailand



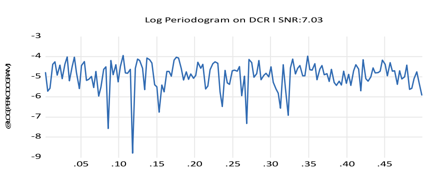
Panel 9 (Periodogram): Philippines



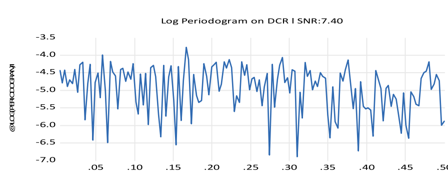
Panel 10 *(Periodogram):* Malaysia



Panel 11 *(Periodogram):* South Korea



Panel 12 *(Periodogram):* Indonesia



stationary series. The highest peaks take place around April 1999 (Thailand, corresponding to frequency 36.5), December 1992 (South Korea), May 2000 (Philippines), May 2002 (India), September 2002 (Mexico), February 2001 (Taiwan), July 2007 (Chile), December 1999 (Colombia), December 1999 (Hong Kong), July 1999 (Malaysia), May 1998 (Indonesia), and May 1999 (Brazil)[14]. Second, we also perform a series of Fisher's g *t*-tests. We find that the hypothesis that the periodogram ordinate is not significant enough cannot be rejected as their *t*-state values are negative for all series in our study. Third, we also find that all the normalized integrated spectrum tests fail to reject the hypothesis that the observations from each series come from a Gaussian white noise process, as the statistics fall within the parallel lines. Thus, it appears that our series are stationary. In a related study by Durlauf (1991), the author shows evidence that the random walk theory is rejected for the US data. To save journal space only the 12 log periodograms are included in the article. The rest are available in the supplemental information.

8. Concluding Remarks

As investments into emerging markets have continually been increasing the past two decades, it has become ever more important to examine the topic of market efficiency for these markets. We set out to explore stock price predictability in a group of 12 emerging markets using a recent data set from the reputed WRDS database and improved statistical methods. We contribute to the existing literature by employing a large, unique data set, across 12 countries, and a comprehensive set of traditional and recent statistical methods as well as powerful multiple-break unit root and spectral analysis tests, many of which have never been used to evaluate the efficient market hypothesis (EMH) in emerging markets. We find that, with a few exceptions, emerging stock markets are generally not efficient. This may be due to more pronounced information asymmetry, timeliness of trading information, tighter government regulations, accuracy of financial data and economic systems not conducive to the free flow of available information to all market participants. These results may also suggest that emerging markets are not fully integrated with world capital markets. Based on our findings, it may be concluded investors could earn arbitrage profits due to market inefficiency in these countries. Further, investors run the risk of not having the necessary information to make fully formed decisions. It should be expected that (once the governing bodies and institution in each country understand and accept their market inefficiencies) governments in those countries would seek to improve control, transparency, information sharing, and integration into world markets so that their markets will perform increasingly more like the developed markets.

Notes

1. The 12 countries are: Brazil, Chile, Columbia, Hong Kong, India, Indonesia, Malaysia, Mexico, Philippines, Taiwan, Thailand. Selection of these countries is discussed in the data section.
2. 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 Stracivich, 2003; Baltagi, 2008).
3. In addition, we looked at using GARCH-style unit root tests. After careful examination of our data, the lack of volatility clustering in our return series and formal tests of conditional heteroskedasticity suggest that this newer method is not appropriate for our data.
4. Our data spans a relatively long period of time using daily frequency. For most countries, the data start from the late 1980s and end in 2015.
5. The correlation coefficients among the WRDS indices and those of COMPUSTAT are between 0.95 and 0.98 for the countries in our sample, according to WRDS.
6. We also performed the single-break Zivot–Andrews test and found that most of the answers are similar to those of the multiple unit root tests. Due to its shortcomings and to converse space, we do not report the results which are available from the authors upon request.
7. To conserve space, we report the results for 2 lags since the results are essentially the same any of these methods.
8. We thank an anonymous subject editor for the suggestion to include this test.

9. We thank an anonymous referee for his/her suggestion to include this analysis into the article, along with other suggestions that improve the quality of our study.
10. We first examine the periodogram for each series and, when relevant, the smoothed (log) version of the data which has the advantage of stabilizing the variance and providing a closer look of the spectrum.
11. Results using other lags (0,4) are qualitatively similar and are thus not reported, to conserve space. Results are also invariant when using the AIC criterion.
12. When employing 3 breaks with the LP test, it should be noted that the critical values for Chile cannot be computed. This is not the case with most other countries.
13. To conserve space, we do not show the regular periodograms as they are strikingly similar. Specifically, they do not suggest the presence of a fixed cycle. Rather, they seem to indicate a stochastic cycle or a pseudo-cyclical pattern, similar to those of an autoregressive moving average model. The log periodograms with significantly higher signal to noise ratios (SNR) – about 3 times larger – provide a much closer look at the series and are shown in Table 6 instead.
14. These values can be seen from the graphical outputs or computed from the data. In many cases, the highest peaks for each series tend to happen in the periods following the Asian Financial Crisis, not surprisingly.

References

Amini, S., R. Hudson, and K. Keasey. 2010. Stock return predictability despite low autocorrelation. *Economics Letters* 108:101–03. doi:10.1016/j.econlet.2010.04.031.

Arltova, A. J. 2000. *Variance ratios, socio-economical applications of statistical methods*. W. Ostasiewicz, Edited by. 17029. Wroclaw, Poland: Wroclaw University of Economics Publishing House

Bekaert, G., and C. R. Harvey. 2002. Research in emerging markets finance: Looking to the future. *Emerging Markets Review* 3:429–48. doi:10.1016/S1566-0141(02)00045-6.

Bertone, S., I. Paeglis, and R. Ravi. 2015. Has the market become more efficient? *Journal of Banking & Finance* 54:72–86. doi:10.1016/j.jbankfin.2014.12.019.

Bhargava, A. 1986. On the theory of testing for unit roots in observed time series. *Review of Economic Studies* 53:369–84. doi:10.2307/2297634.

Borges, M. 2010. Efficient market hypothesis in European stock markets. *European Journal of Finance* 16:711–26. doi:10.1080/1351847X.2010.495477.

Borges, M. 2011. Random walk tests for the Lisbon stock market. *Applied Economics* 43:631–39. doi:10.1080/00036840802584935.

Brock, W. A., W. D. Deschert, J. A. Scheinkman, and B. LeBaron. 1996. (revised version) A test for independence based on the correlation dimension. *Econometric Reviews* 15:197–235.

Busse, J. A., A. Goyal, and S. Wahal. 2010. Performance and persistence in institutional investment management. *Journal of Finance* 65:765–90. doi:10.1111/j.1540-6261.2009.01550.x.

Carhart, M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52:57–82. doi:10.1111/j.1540-6261.1997.tb03808.x.

Charles, A., and O. Darné. 2009. Variance ratio tests of random walk: An overview. *Journal of Economic Surveys* 23:503–27. doi:10.1111/j.1467-6419.2008.00570.x.

Chow, K. V., and K. C. Denning. 1993. A simple multiple variance ratio test. *Journal of Econometrics* 58:385–401. doi:10.1016/0304-4076(93)90051-6.

De Lima, P. F. 1998. Nonlinearities and Nonstationarities in Stock Returns. *Journal of Business and Economic Statistics* 16 (2):227–36.

De Long, J. B., and M. Becht (1992). Excess volatility and the German Stock Market, NBER working papers 133, University of Califonia at Berkeley, Economics Department.

Durlauf, S. 1991. Spectral based testing of the martingale hypothesis. *Journal of Econometrics* 50:355–76. doi:10.1016/0304-4076(91)90025-9.

Elliot, G., T. Rothernberg, and J. Stock. 1996. Efficient tests for an autoregressive unit root. *Econometrica* 64:831–36.

Enders, W. 2008. *Applied econometric time series*. Hoboken (NJ): Wiley.

Fama, E. F., and K. R. French. 2010. Luck versus skill in the cross section of mutual fund returns. *Journal of Finance* 65:1915–47. doi:10.1111/j.1540-6261.2010.01598.x.

French, K. R., and R. Roll. 1986. The arrival of information and the reaction of traders. *Journal of Financial Economics* 17 (1):5–26. doi:10.1016/0304-405X(86)90004-8.

Griffin, J. M., P. J. Kelly, and F. Nardari. 2010. Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. *Review of Financial Studies* 23:3225–77. doi:10.1093/rfs/hhq044.

Grossman, S., and J. Stiglitz. 1976. Information and competitive price systems. *American Economic Review* 66:246–52.

Hadhri, S., and Z. Ftiti. 2017. Stock return predictability in emerging markets: Does the choice of predictors and models matter across countries. *Research in International Business and Finance* 42:39–60. doi:10.1016/j.ribaf.2017.04.057.

Henry, P. B. 2000. Stock Market Liberalization, Economic Reform, and emerging Market Equity Prices. *Journal of Finance* 55:529–64. doi:10.1111/0022-1082.00219.

Hiremath, G. S., and J. Kumari (2014). Stock returns predictability and the adaptive market hypothesis in emerging markets: Evidence from India. Munich Personal RePEc Archive Paper 58378.

Kaminsky, G., and S. Schmukler (2002). Short-Run Pain, Long-Run Gain: The effects of financial liberalization. Policy Research Working Papers, World Bank October 2002.

Kheradyar, S., I. Ibrahim, and F. Mat Nor. 2011. Stock return predictability with financial ratios. *International Journal of Trade, Economics and Finance* 2 (5):391–96. doi:10.7763/IJTEF.2011.V2.137.

Lee, J., and M. C. Strazicich. 2003. Minimum Lagrange multiplier unit root tests with two structural breaks. *Review of Economics and Statistics* 85:1082–89. doi:10.1162/003465303772815961.

Lettau, M., and S. Van Nieuwerburgh. 2008. Reconciling the return predictability evidence. *Review of Financial Studies* 21:1607–52. doi:10.1093/rfs/hhm074.

Levin, A., C. Lin, and C. J. Chu. 2002. Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics* 108 (1):1–24. doi:10.1016/S0304-4076(01)00098-7.

Lim, K. P., and R. D. Brooks (2006). The evolving and relative efficiencies of stock markets: Empirical evidence from rolling bicorrelation test statistics. SSRN. September 2006.

Lim, K. P., W. Luo, and J. H. Kim. 2013. Are US stock index returns predictable? Evidence from automatic autocorrelation-based tests. *Applied Economics* 45 (8):953–62. doi:10.1080/00036846.2011.613782.

Lo, A. 2004. The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management* 20 (5):15–29. doi:10.3905/jpm.2004.442611.

Lo, A. W., and A. C. MacKinlay. 1988. Stock market prices do not follow random walks. *Evidence from a Simple Specification Test, Review of Financial Studies* 1:41–66. doi:10.1093/rfs/1.1.41.

Lumsdaine, R., and D. Papell. 1997. Multiple trend breaks and the unit-root hypothesis. *Review of Economics and Statistics* 79:212–18. doi:10.1162/003465397556791.

Malafeyev, O., A. Awasthi, and K. S. Kambekar (2017). Random walks and market efficiency in Chinese and Indian equity markets. Papers 1709.04059, found at, www.arXiv.org Jan 20 2018.

Mitra, S. K., M. Chattopadhyay, P. Charan, and J. Bawa. 2017. Identifying periods of market inefficiency for return predictability. *Applied Economics Letters* 24 (10):668–71. doi:10.1080/13504851.2016.1218424.

Narayan, P. K., and S. Popp. 2010. A new unit root test with two structural breaks in level and slope at unknown time. *Journal of Applied Statistics* 37 (9):1425–38. doi:10.1080/02664760903039883.

Narayan, P. K., and S. Popp. 2013. Size and power properties of structural break unit root tests. *Applied Economics* 45 (6):721–28. doi:10.1080/00036846.2011.610752.

Ng, S., and P. Perron. 2001. Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69 (6):1519–54. doi:10.1111/1468-0262.00256.

Perron, P., and S. NG. 1996. Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *The Review of Economic Studies* 62 (3):435–63. doi:10.2307/2297890.

Phillips, P., and P. Perron. 1988. Testing for a unit root in time series regressions. *Biometrika* 75 (2):335–46. doi:10.1093/biomet/75.2.335.

Schwert, G. W. 1989. Why does Stock Market Volatility Change Over Time? *Journal of Finance* 44 (5):1115–11153. doi:10.1111/j.1540-6261.1989.tb02647.x.

Seth, N., and K. Sharma. 2015. International stock market efficiency and integration: Evidences from Asian and US markets". *Journal of Advances in Management Research* 12 (2):88–106. doi:10.1108/JAMR-07-2011-0010.

Shiller, R. 1981. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review, American Economic Association* 71 (3):421–36.

Smith, G. 2011. The changing and relative efficiency of European emerging stock markets. *European Journal of Finance* 8:689–708.

Stiglitz, J. (2014). Tapping the Brakes: Are less active markets safer and better for the economy? Federal Reserve Bank Atlanta Presentation.

Urquhart, A., and F. McGroarty. 2016. Are stock markets really efficient? Evidence of the adaptive market hypothesis. *International Review of Financial Analysis* 47:39–49. doi:10.1016/j.irfa.2016.06.011.

Urquhart, A., and R. 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. doi:10.1016/j.irfa.2013.03.005.

Verheyden, T., L. Moor, and F. Bossche. 2015. Towards a new framework on efficient markets. *Research in International Business and Finance* 34:294–308. doi:10.1016/j.ribaf.2015.02.007.

Warner, R. 1998. *Spectral analysis of time series data*. NYC, New York: Guilford Press.

Wei, W. S. 2018. *Time series analysis, Second edition*. Boston (MA): Addison Wesley.

Wright, J. 2000. Alternative variance-ratio tests using ranks and signs. *Journal of Business and Economic Statistics* 18:1–9.

Xiong, X., Y. Meng, X. Li, and D. Shen. 2019. An empirical analysis of the adaptive market hypothesis with calendar effects: Evidence from China. *Finance Research Letters* 31 (c). doi:10.1016/j.frl.2018.11.020.