

Testing the white noise hypothesis of stock returns<sup>☆</sup>Jonathan B. Hill<sup>a</sup>, Kaiji Motegi<sup>b,\*</sup><sup>a</sup> Department of Economics, University of North Carolina, Chapel Hill, United States<sup>b</sup> Graduate School of Economics, Kobe University, 2-1 Rokkodai-cho, Nada, Kobe, Hyogo 657-8501, Japan

## ARTICLE INFO

## JEL classifications:

C12  
C58  
G14

## Keywords:

Blockwise wild bootstrap  
Randomized block size  
Serial correlation  
Weak form efficiency  
White noise test

## ABSTRACT

Weak form efficiency of stock markets implies unpredictability of stock returns in a time series sense, and the latter is tested predominantly under a serial independence or martingale difference assumption. Since these properties rule out weak dependence that may exist in stock returns, it is of interest to test whether returns are white noise. We perform white noise tests assisted by Shao's (2011) blockwise wild bootstrap. We reveal that, in rolling windows, the block structure inscribes an artificial periodicity in bootstrapped confidence bands. We eliminate the periodicity by randomizing a block size. The white noise hypothesis is accepted for Chinese and Japanese markets, suggesting that those markets are weak form efficient. The white noise hypothesis is rejected for U.K. and U.S. markets during the Iraq War and the subprime mortgage crisis due to significantly negative autocorrelations, suggesting that those markets are inefficient in crisis periods.

## 1. Introduction

We test for stock market efficiency over rolling data sample windows, and make new contributions to the study of white noise tests. A stock market is *weak form efficient* if stock prices fully reflect historical price information (cf. Fama, 1970). Weak form efficiency implies unpredictability of stock returns in a time series sense, and the latter is extensively tested via various methods. Empirical results have been mixed, with substantial debate between advocates of the efficient market hypothesis (EMH) and proponents of behavioral finance.<sup>1</sup> More recently, the adaptive market hypothesis (AMH) proposed by Lo (2004, 2005) attempts to reconcile the two opposing schools, arguing that the degree of stock market efficiency varies over time.

In line with those trends, a number of recent applications perform rolling window analysis in order to investigate the dynamic evolution of stock market efficiency. See Kim and Shamsuddin (2008),

Lim et al. (2008), Kim et al. (2011), Lim et al. (2013), Verheyden, De Moor, and Van den Bossche (2015), Anagnostidis et al. (2016), Urquhart and McGroarty (2016), Charles et al. (2017), and Mitra et al. (2017) for recent studies. An advantage of rolling window analysis relative to non-overlapping subsample analysis is that the former does not require a subjective choice of the first and last dates of major events such as financial crises.

A common perception in the literature is that a financial crisis augments investor panic and thus lowers the degree of market efficiency. Lim, Brooks, and Kim (2008) and Anagnostidis et al. (2016) find that market efficiency is indeed adversely affected by financial crises. By comparison, Kim and Shamsuddin (2008), Verheyden, De Moor, and Van den Bossche (2015), and Charles et al. (2017) find mixed results.

In terms of methodology, many early contributions use Lo and MacKinlay's (1988) variance ratio test and its variants.<sup>2</sup> Other tests include Hong's (1999) spectral test, Wright's (2000) non-parametric sign/rank test, Escanciano and Valasco's (2006) generalized spectral

<sup>☆</sup> We thank a referee, an associate editor, and co-editor Sushanta Mallick for helpful comments. We also thank Eric Ghysels, Shigeyuki Hamori, Peter R. Hansen, Yoshihiko Nishiyama, and Ke Zhu, seminar participants at Kyoto University, the University of North Carolina at Chapel Hill, and the University of Hong Kong, and conference participants at the 11th Spring Meeting of the Japan Statistical Society, EcoSta 2017, 4th Annual Conference of IAAE, and 3rd Annual International Conference on Applied Econometrics in Hawaii for helpful comments and discussions. The second author is grateful for financial supports from JSPS KAKENHI (Grant Number 16K17104), Kikawada Foundation, Mitsubishi UFJ Trust Scholarship Foundation, Nomura Foundation, and Suntory Foundation.

\* Corresponding author.

E-mail addresses: [jhill@email.unc.edu](mailto:jhill@email.unc.edu) (J.B. Hill), [motegi@econ.kobe-u.ac.jp](mailto:motegi@econ.kobe-u.ac.jp) (K. Motegi).

<sup>1</sup> See Yen and Lee (2008) and Lim and Brooks (2009) for extensive surveys of stock market efficiency.

<sup>2</sup> See Charles and Darné (2009) for a survey of variance ratio tests.

test, and Escanciano and Lobato's (2009) robust automatic portmanteau test.

Note that the implicit null hypothesis of all tests above is either that returns are serially independent or a martingale difference sequence (mds), since the utilized asymptotic theory requires such assumptions under the null hypothesis. These properties rule out higher forms of dependence that may exist in stock returns, while the mds property is generally not sufficient for a Gaussian central limit theory (e.g. Billingsley, 1961). Moreover, their test is not a true white noise test since it does not take into account serial correlations at *all* lags asymptotically.

A promising alternative is a white noise test with only serial uncorrelatedness under the null hypothesis, as well as standard higher moment and weak dependence properties to push through standard asymptotics. A rejection of the white noise hypothesis is supposed to serve as a helpful signal for arbitragers, since a rejection indicates the existence of non-zero autocorrelation at some lags. We place the present study in the literature that is strongly interested in whether asset returns are white noise, a useful albeit weak measure of market efficiency.<sup>3</sup>

Formal white noise tests with little more than serial uncorrelatedness under the null have not been available until recently. See Hill and Motegi (2017) for many detailed references, some of which are discussed below. Conventional portmanteau tests bound the maximum lag and therefore are not true white noise tests, although weak dependence, automatic lag selection, and a pivotal structure irrespective of model filter are allowed.<sup>4</sup> Hong (1996, 2001) standardizes a portmanteau statistic, allowing for an increasing number of serial correlations and standard asymptotics.

Spectral tests operate on the maximum (and therefore increasing) number of serial correlations as the sample size increases, with variations due to Durlauf (1991), Hong (1996), Deo (2000), and Delgado et al. (2005). Cramér-von Mises and Kolmogorov-Smirnov variants can be found in Shao (2011) with a blockwise wild bootstrap procedure that allows for weak dependence under the null. A weighted sum of serial correlations also arises in the white noise test of Andrews and Ploberger (1996) (cf. Nankervis and Savin, 2010).

Hill and Motegi (2017) develop a new theory of white noise tests. They allow for a very broad class of dependent and heterogeneous data, and verify that Shao's (2011) blockwise wild bootstrap is valid in such a general setting. Based on their theory, this paper uses the Cramér-von Mises test assisted by Shao's (2011) blockwise wild bootstrap to test if stock returns are white noise.

We analyze daily stock price indices from China, Japan, the U.K., and the U.S. The entire sample period spans January 2003 through October 2015. We perform a rolling window analysis in order to capture time-varying market efficiency. The degree of stock market efficiency is expected to be time-dependent in view of the literature of the adaptive market hypothesis. It is of particular interest to see how market efficiency is affected by financial turbulence like the subprime mortgage crisis around 2008.

We are not aware of any applications of the blockwise wild bootstrap, except for Shao (2010, 2011) who analyzes temperature data and stock returns in a full sample framework. The present study is therefore the first use of the blockwise wild bootstrap in a rolling window framework, in which we found and corrected a key shortcoming. In rolling window sub-samples, the block structure inscribes an artificial periodicity in the bootstrapped data, and therefore in confidence bands. A similar periodicity occurs in the block bootstrap for dependent data, which Politis and Romano (1994) correct by randomizing a block size.

<sup>3</sup> Another interpretation of the present analysis is testing if stock prices follow what Campbell et al. (1997) call Random Walk 3 (i.e. random walk with uncorrelated increment) in a strict sense.

<sup>4</sup> See Romano and Thombs (1996), Lobato (2001), Lobato et al. (2002), Horowitz et al. (2006), Mantalos and Shukur (2008), Escanciano and Lobato (2009), Delgado and Velasco (2011), Guay et al. (2013), Zhu and Li (2015), and Zhang (2016).

We take the same approach to eliminate blockwise wild bootstrap periodicity. See Section 2 for key details, and see the supplemental material Hill and Motegi (2018, Sections 3–6) for complete details.

We find that the degree of market efficiency varies across countries and sample periods, as asserted by the adaptive market hypothesis. For Chinese and Japanese stock markets, the white noise hypothesis is accepted in most rolling windows, which is evidence for weak form efficiency. The same goes for U.K. and U.S. markets during non-crisis periods. When the U.K. and U.S. markets face greater uncertainty, namely the Iraq War and the subprime mortgage crisis, the white noise hypothesis is often rejected, which is evidence against weak form efficiency. Our empirical results are consistent with Kim and Shamsuddin (2008), Verheyden, De Moor, and Van den Bossche (2015), and Charles et al. (2017) in the sense that a financial crisis lowers the degree of market efficiency in some countries but not in others.

Importantly, the rejection of white noise hypothesis in the U.K. and U.S. markets during crisis periods stems from large negative autocorrelations at lag 1. A negative correlation, in particular at low lags, signifies rapid changes in market trading, which is corroborated with high volatility during these times. The appearance of negative autocorrelations in short (e.g. daily or weekly) and long (e.g. 1 or 3-year) horizon returns is well documented. Evidence for positive or negative correlations depends heavily on the market, return horizon, and the presence of crisis periods. See, e.g., Fama and French (1988), who argue that predictable price variation due to mean-reversion in returns accounts for the negative correlation at short and long horizons.

The remainder of the paper is organized as follows. In Section 2 we explain the white noise tests and the blockwise wild bootstrap. Section 3 describes our data, and we report the empirical results in Section 4. Concluding remarks are provided in Section 5. A supplemental analysis using aggregated weekly and monthly data is performed in Appendix. See the supplemental material Hill and Motegi (2018) for omitted details on theory, simulations, and empirical analysis.

## 2. Methodology

### 2.1. White noise test

Let  $P_t$  be a stock price index at day  $t \in \{1, 2, \dots, n\}$ , then  $r_t = \ln(P_t/P_{t-1})$  is the log return. Define population mean  $\mu = E[r_t]$ , variance  $\gamma(0) = E[(r_t - \mu)^2]$ , autocovariance  $\gamma(h) = E[(r_t - \mu)(r_{t-h} - \mu)]$ , and autocorrelation  $\rho(h) = \gamma(h)/\gamma(0)$  for  $h \geq 1$ . We wish to test for the white noise hypothesis of the log return  $r_t$ :

$$H_0 : \rho(h) = 0 \quad \text{for all } h \geq 1 \text{ against } H_1 : \rho(h) \neq 0 \text{ for some } h \geq 1.$$

A rejection of  $H_0$  is evidence against weak form efficiency of the stock market, while a non-rejection is evidence for weak form efficiency.

Define sample mean  $\hat{\mu}_n = (1/n) \sum_{t=1}^n r_t$ , variance  $\hat{\gamma}_n(0) = (1/n) \sum_{t=1}^n (r_t - \hat{\mu}_n)^2$ , autocovariance  $\hat{\gamma}_n(h) = (1/n) \sum_{t=h+1}^n (r_t - \hat{\mu}_n)(r_{t-h} - \hat{\mu}_n)$ , and autocorrelation  $\hat{\rho}_n(h) = \hat{\gamma}_n(h)/\hat{\gamma}_n(0)$  for  $h \geq 1$ . In this paper we use the Cramér-von Mises [CvM] statistic, which is based on the sample spectral density (cf. Shao, 2011):

$$C_n = n \int_0^\pi \left\{ \sum_{h=1}^{n-1} \hat{\gamma}_n(h) \psi_h(\lambda) \right\}^2 d\lambda, \quad \text{where } \psi_h(\lambda) = (h\pi)^{-1} \sin(h\lambda). \quad (1)$$

By construction, all  $n - 1$  possible lags are used, and asymptotically  $\gamma(h)$  is estimated for every integer  $h \geq 1$ . Thus, the CvM statistic can be used for a formal white noise test. The test statistic has a non-standard limit distribution under  $H_0$ , and Shao (2011) proves that the blockwise wild bootstrap, which is a version of the dependent wild bootstrap proposed in Shao (2010), is valid under certain conditions on moments

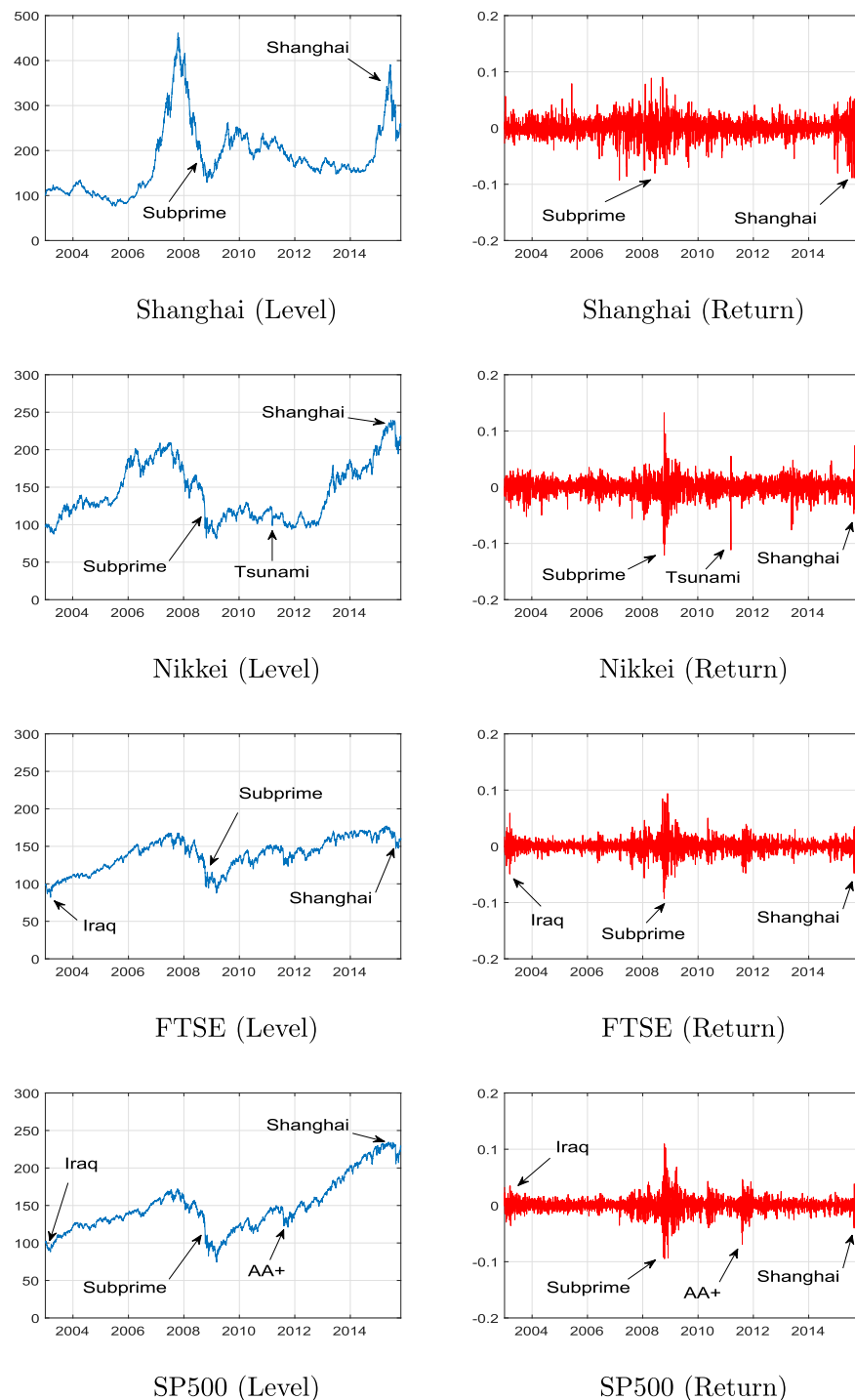


Fig. 1. Daily stock prices and log returns (01/01/2003–10/29/2015).

and dependence. Hill and Motegi (2017) further relaxed the conditions imposed by Shao (2011). We therefore use the blockwise wild bootstrap to perform the CvM test.<sup>5</sup>

<sup>5</sup> In the supplemental material Hill and Motegi (2018, Section 7.2), we also perform Hill and Motegi's (2017) max-correlation test and Andrews and Ploberger's (1996) sup-LM test. As proven by Hill and Motegi (2017), those tests are asymptotically valid when assisted by the blockwise wild bootstrap. Hill and Motegi (2017) also find that those tests are comparable with the CvM test in terms of empirical size and power.

## 2.2. Blockwise wild bootstrap

The blockwise wild bootstrap is executed as follows. Set a block size  $b_n$  such that  $1 \leq b_n < n$ . Generate iid random numbers  $\{\xi_1, \dots, \xi_{n/b_n}\}$  with  $E[\xi_i] = 0$ ,  $E[\xi_i^2] = 1$ , and  $E[\xi_i^4] < \infty$ . Assume for simplicity that the number of blocks  $n/b_n$  is an integer. Standard normal  $\xi_i$  satisfies these properties, and is used in the empirical analysis below. Define an auxiliary variable  $\omega_t$  block-wise as follows:  $\{\omega_1, \dots, \omega_{b_n}\} = \xi_1$ ,  $\{\omega_{b_n+1}, \dots, \omega_{2b_n}\} = \xi_2$ , ...,  $\{\omega_{(n/b_n-1)b_n+1}, \dots, \omega_n\} = \xi_{n/b_n}$ . Thus,  $\omega_t$  is iid across blocks, but perfectly dependent within blocks. Compute a

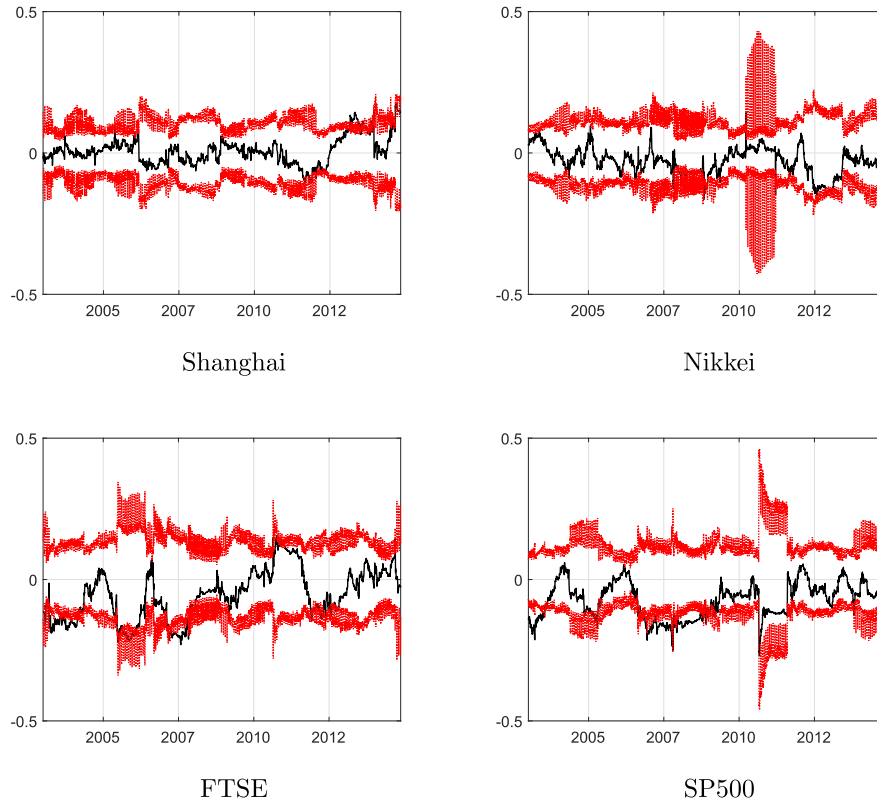


Fig. 2. Autocorrelations at Lag 1 with Window Size  $n = 240$  (Fixed Block Size).

bootstrapped autocovariance

$$\hat{\gamma}_n^{(bw)}(h) = \frac{1}{n} \sum_{t=h+1}^n \omega_t \{ (r_t - \hat{\mu}_n)(r_{t-h} - \hat{\mu}_n) - \hat{\gamma}_n(h) \} \text{ for } h = 1, \dots, n-1, \quad (2)$$

and a bootstrapped CvM test statistic  $\hat{C}_n^{(bw)}$  by replacing  $\hat{\gamma}_n(h)$  in (1) with  $\hat{\gamma}_n^{(bw)}(h)$ . Repeat  $M$  times, resulting in  $\{\hat{C}_{n,i}^{(bw)}\}_{i=1}^M$ . The approximate p-value is  $\hat{p}_{n,M}^{(bw)} \equiv (1/M) \sum_{i=1}^M I(\hat{C}_{n,i}^{(bw)} \geq \hat{C}_n^{(bw)})$ . Now let the number of bootstrap samples satisfy  $M = M_n \rightarrow \infty$  as  $n \rightarrow \infty$ . If  $\hat{p}_{n,M}^{(bw)} < \alpha$ , then reject the null hypothesis of white noise at significance level  $\alpha$ ; otherwise do not reject the null.

The blockwise wild bootstrapped CvM test is asymptotically valid and consistent for a large class of processes that may be dependent under  $H_0$ : the asymptotic probability of rejection at level  $\alpha$  is exactly  $\alpha$  under  $H_0$ , and it is equal to 1 under  $H_1$  (see Shao, 2011; Hill and Motegi, 2017).<sup>6</sup>

In order to test  $\rho(h) = 0$  for a specific horizon  $h$  or any set of horizons, (2) can be used to construct a bootstrapped correlation  $\hat{\rho}_n^{(bw)}(h) = \hat{\gamma}_n^{(bw)}(h)/\hat{\gamma}_n(0)$  and a confidence band under  $\rho(h) = 0$ . Compute  $\{\hat{\rho}_{n,i}^{(bw)}(h)\}_{i=1}^M$  and sort as  $\hat{\rho}_{n,[1]}^{(bw)}(h) \leq \hat{\rho}_{n,[2]}^{(bw)}(h) \leq \dots \leq \hat{\rho}_{n,[M]}^{(bw)}(h)$ . The 95% band for lag  $h$  is then  $[\hat{\rho}_{n,[0.025 \times M]}^{(bw)}(h), \hat{\rho}_{n,[0.975 \times M]}^{(bw)}(h)]$ . In Section 4.1, we compute the bootstrapped 95% confidence band with  $h = 1$ .

<sup>6</sup> Besides the blockwise wild bootstrap, Zhu and Li's (2015) block-wise random weighting bootstrap can be applied to the CvM test statistic. Hill and Motegi (2017) find that both bootstrap procedures are comparable in terms of empirical size and power.

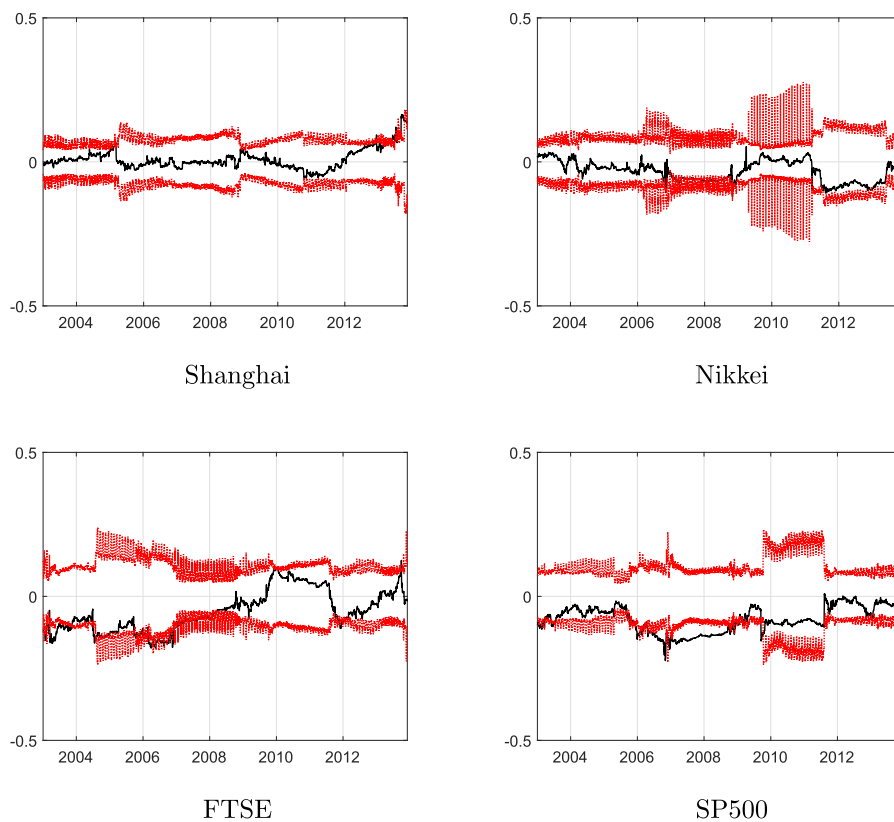
### 2.3. Periodic confidence bands over rolling windows

The blockwise wild bootstrap has not been studied in a rolling window environment. Using the same auxiliary variable  $\omega_t$  in each block of size  $b_n \rightarrow \infty$ , with  $b_n = o(n)$ , is key toward allowing for general dependence under the null. Unfortunately, in a *rolling window* setting, the consequence is a periodically fluctuating p-value or confidence band, irrespective of the true data generating process (e.g. periodic fluctuations arise even for iid data).

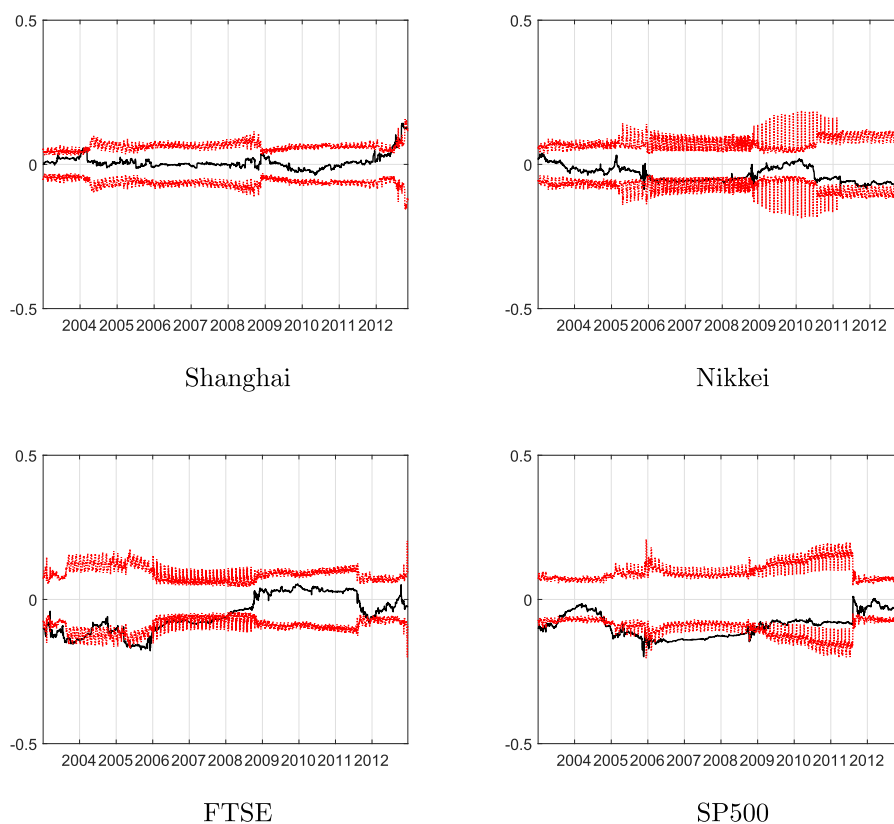
The reason for the artificial periodicity is that similar blocking structures keep arising every  $b_n$  windows. Consider two windows that are apart from each other by  $b_n$  windows (e.g. the first window is  $\{1, 2, \dots, n\}$  and the  $(b_n + 1)$ -th window is  $\{b_n + 1, b_n + 2, \dots, b_n + n\}$ ), where  $n$  is the size of a window. A block in one window is a scalar multiplication of a block in the other window, resulting in similar bootstrapped autocorrelations from those two windows. See the supplemental material Hill and Motegi (2018, Sections 3–6) for complete details and additional simulations that demonstrate the problem and solution. The problem is well known in the context of the block bootstrap (e.g. Politis and Romano, 1994), but has been ignored for the blockwise wild bootstrap.

As in Politis and Romano (1994) (cf. Lahiri, 1999), we resolve the problem by randomizing the block size for each bootstrap sample and window. Randomness across windows ensures that different windows have different blocking structures, and are therefore not multiples of each other. This removes the artificial periodicity successfully. Randomness across bootstrap samples makes the confidence bands less volatile, which is desired in terms of visual inspection.

In a full sample environment, Shao (2011) finds that block sizes  $b_n = c\sqrt{n}$  with  $c \in \{0.5, 1, 2\}$  perform comparably. We therefore pick a middle value  $c = 1$  and add randomness as follows. We draw a uniform random variable  $c$  on  $[1 - \iota, 1 + \iota]$  with  $\iota = 0.5$  for each boot-



**Fig. 3.** Autocorrelations at Lag 1 with Window Size  $n = 480$  (Fixed Block Size).



**Fig. 4.** Autocorrelations at Lag 1 with Window Size  $n = 720$  (Fixed Block Size).

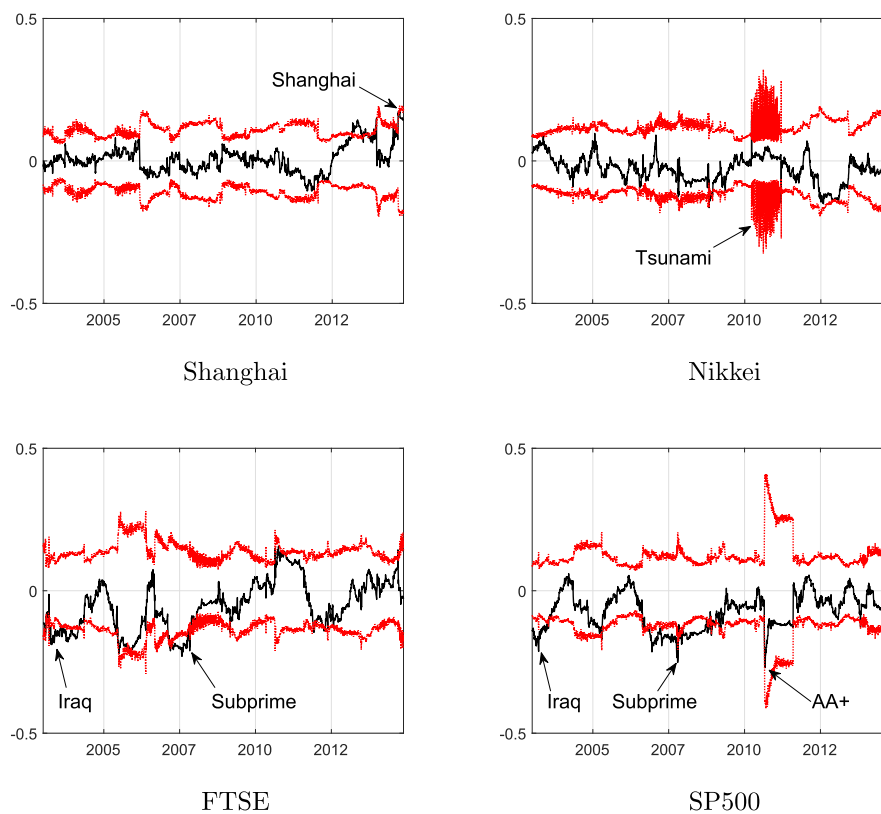


Fig. 5. Autocorrelations at Lag 1 with Window Size  $n = 240$  (Randomized Block Size).

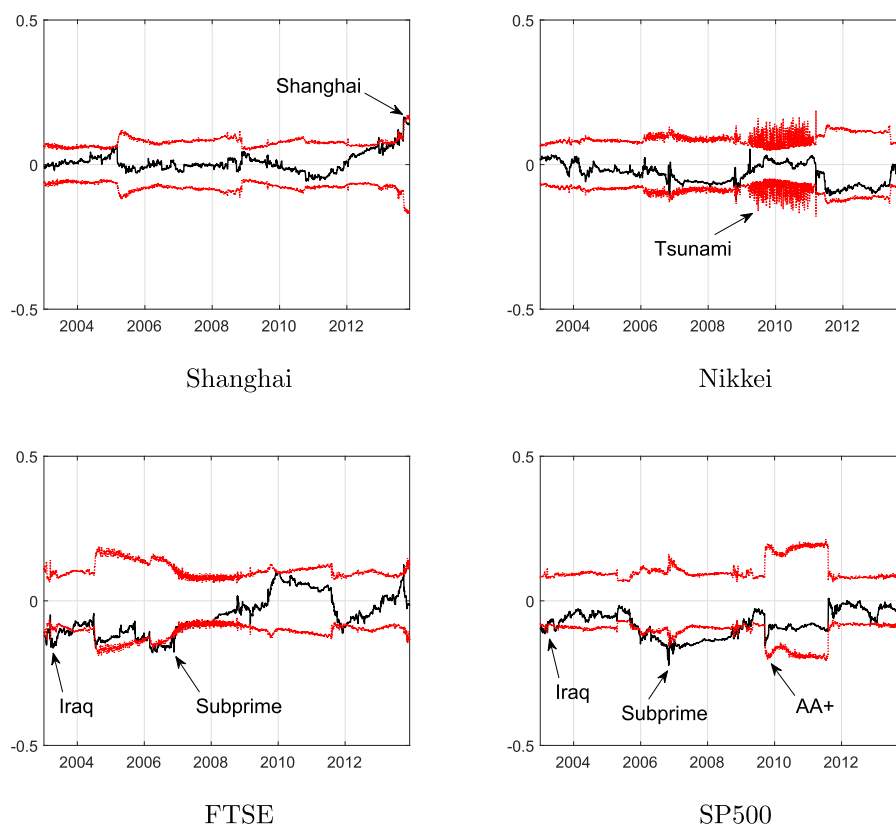


Fig. 6. Autocorrelations at Lag 1 with Window Size  $n = 480$  (Randomized Block Size).



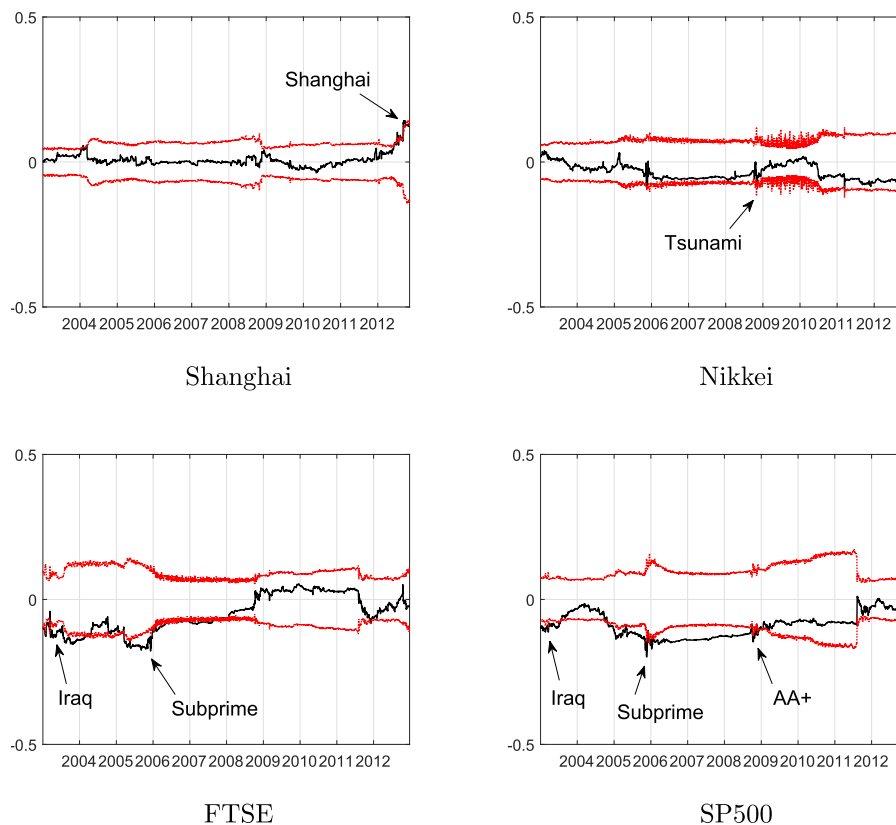


Fig. 7. Autocorrelations at Lag 1 with Window Size  $n = 720$  (Randomized Block Size).

strap sample and window, and use  $b_n = c\sqrt{n}$ .<sup>7</sup>

### 3. Data

We analyze log returns of the daily closing values of the Shanghai Stock Exchange Composite Index in China (called “Shanghai” hereafter), the Nikkei 225 Index in Japan (“Nikkei”), the FTSE 100 Index in the U.K. (“FTSE”), and the S&P 500 Index in the U.S. (“SP500”), all in local currencies from January 1, 2003 through October 29, 2015.<sup>8</sup> The Shanghai index is selected as a representative of emerging markets, while Nikkei, FTSE, and SP500 are selected as representatives of some of the most liquid, mature, and influential markets. For each series, we use a price return index as opposed to a total return index, hence we do not take dividends into account. We choose the price return index since it allows us to have long enough sample periods. Sample size differs across countries due to different holidays and other market closures: 3110 days for Shanghai, 3149 for Nikkei, 3243 for FTSE, and 3230 for SP500.<sup>9</sup>

Fig. 1 plots each stock price index and log return. For the former, the initial value is normalized at 100 for visual clarity. The subprime mortgage crisis in 2007–2008 (annotated as “Subprime” in the figure)

<sup>7</sup> It remains an open question how to pick  $\iota$  in practice. There is a trade-off such that a small  $\iota$  does not fully eliminate periodicity due to little randomness, while a large  $\iota$  results in more volatility in confidence bands. In this paper we simply choose  $\iota = 0.5$  as a rule of thumb, and verify via controlled experiments and empirical analysis that our choice yields sufficiently non-periodic, smooth confidence bands.

<sup>8</sup> The data were retrieved from Bloomberg.

<sup>9</sup> Market closures are simply ignored: log return on day  $t$  is computed as the log difference between a price on day  $t$  and a price on the previous business day.

caused a dramatic decline in the stock prices. FTSE and SP500 experienced relatively fast recovery from the stock price plummet in 2008, while Shanghai and Nikkei experienced a longer period of stagnation. Each return series shows clear volatility clustering, especially during the subprime mortgage crisis.

Besides the subprime mortgage crisis, there are a few episodes of financial turbulence that may affect the autocorrelation structure of each market. First, in 2014–2015, the Shanghai stock market experienced an apparent “bubble and burst”, and its magnitude is only slightly smaller than the subprime mortgage crisis (annotated as “Shanghai”). Second, Nikkei had a large negative log return of  $-0.112$  on March 15, 2011, which is two business days after the Great East Japan Earthquake. Nuclear power plants in Fukushima were destroyed by a resulting tsunami, and investors had pessimistic sentiment on electricity supply (“Tsunami”). Third, in 2002 and 2003, the FTSE stock market faced a period of great uncertainty due to an economic recession, soaring oil prices, and the Iraq War (“Iraq”). FTSE fell for nine consecutive trading days in January 2003, losing 12.4% of its value. On March 13, 2003, FTSE had a rebound log return of 0.059, indicating an unstable market condition. Fourth, SP500 generated a log return of  $-0.069$  on August 8, 2011 shortly after Standard & Poor’s downgraded the federal government credit rating from AAA to AA+ (“AA+”).

Each log return series has a positive mean, but it is not significant at the 5% level according to a bootstrapped confidence band. Shanghai returns have the largest standard deviation of 0.017, but Nikkei returns have the greatest range of  $[-0.121, 0.132]$ : it has the largest minimum and maximum in absolute value. Due to negative skewness and excess kurtosis, the p-values of the Kolmogorov-Smirnov and Anderson-Darling tests of normality are well below 1% for all markets, strong evidence against normality. See the supplemental material Hill and Motegi (2018, Table 7) for detailed summary statistics.

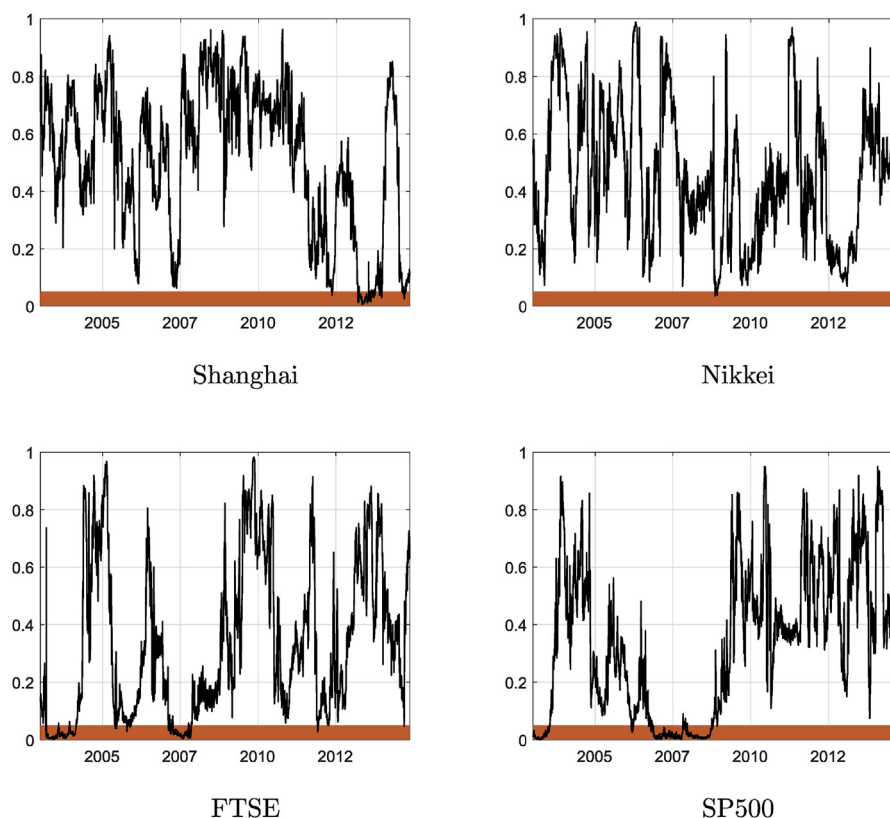


Fig. 8. P-Values of CvM Test with Window Size  $n = 240$  (Randomized Block Size).

We perform the Phillips and Perron (1988) test on the level and log return of the stock price indices. We fail to reject the unit root null hypothesis at any conventional level for the level series, while we reject the unit root null at the 1% level for the log returns. The same result appears regardless of the model specification (i.e. without a constant, with a constant, and with a constant and linear time trend).<sup>10</sup>

#### 4. Empirical results

We perform the rolling window analysis with various window sizes of  $n \in \{240, 480, 720\}$  days (roughly 1, 2, and 3 years). This allows us to see whether results are specific to a window size choice. The first window contains  $\{r_1, \dots, r_n\}$ , the second window contains  $\{r_2, \dots, r_{n+1}\}$ , and the last window contains  $\{r_{T-n+1}, \dots, r_T\}$ , where  $T$  is the entire sample size that differs across countries (see Section 3).

##### 4.1. Analysis of autocorrelation at lag 1

Figs. 2–4 depict sample autocorrelations at lag  $h = 1$  over rolling windows. For each window, the 95% confidence band based on the blockwise wild bootstrap is constructed under the null hypothesis of white noise. The sample correlations are plotted with solid black lines, while the confidence bands are plotted with dotted red lines. We use Shao's (2011) fixed middle block sizes  $b_n = \lfloor \sqrt{n} \rfloor$  so that  $b_n \in \{15, 21, 26\}$  in Figs. 2–4, respectively. The number of bootstrap samples is  $M = 10,000$  for each window. Each point on the horizontal axis

represents the initial date of each window.

A striking result from Figs. 2–4 is that the confidence bands exhibit periodic fluctuations, with the appearance of veritable seasonal highs and lows. Moreover, the zigzag movement repeats itself in every  $b_n$  windows. The confidence bands are particularly volatile for Nikkei and SP500 in 2011, reflecting the tsunami disaster and the securities down-grade shock. Note, however, that periodic confidence bands appear in all series and periods universally.

As discussed in Section 2.3, the periodicity arises because there are similar blocking structures in every  $b_n$  windows.<sup>11</sup> Our proposed solution is randomizing the block size  $b_n$ . For each bootstrap sample, and each window, we independently draw  $c$  from a uniform distribution on  $[0.5, 1.5]$ , and use  $b_n = c\sqrt{n}$ .<sup>12</sup>

See Figs. 5–7 for results with the randomized block size. The periodic fluctuations are successfully removed. We still observe volatile bands for Nikkei in response to the tsunami disaster and for SP500 in response to the U.S. securities downgrade. These are not surprising results since the standard deviation, autocorrelation, and kurtosis of the log returns of those markets all spike in those periods.

In what follows, we discuss the empirical results with the randomized block size and  $n = 240$  (Fig. 5). The larger window sizes yield qualitatively similar implications, although a larger window size naturally leads to less detectable volatility (e.g. the confidence

<sup>11</sup> See the supplemental material Hill and Motegi (2018, Sections 3–6) for computational details and magnified plots.

<sup>12</sup> If we only randomize  $b_n$  across windows (using the same randomized  $b_n$  across all  $M$  bootstrap samples), then the resulting confidence bands have volatility that largely exceeds the volatility of the observed data. By randomizing  $b_n$  across bootstrap draws and windows, both artificial periodicity and excess volatility are eliminated.

<sup>10</sup> The Phillips-Perron test requires a nonparametric variance estimator. We used a Bartlett kernel variance estimator with Newey and West's (1994) automatic lag selection. P-values were computed via MacKinnon's (1996) (one-sided) p-values.



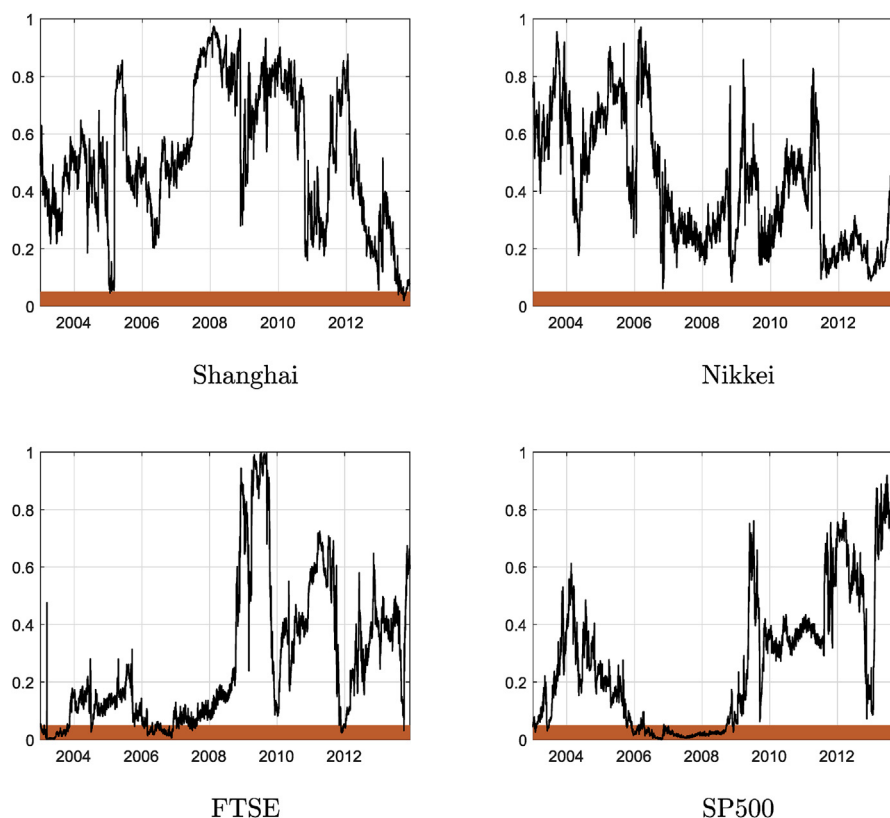


Fig. 9. P-Values of CvM Test with Window Size  $n = 480$  (Randomized Block Size).

bands are monotonically less volatility as the window size increases). For Shanghai, the confidence bands are roughly  $[-0.1, 0.1]$  and they contain the sample correlation in most windows, suggesting that the Shanghai market is weak form efficient. Interestingly, the subprime mortgage crisis around 2008 did not have any noticeable impact on the correlation structure of the Shanghai market, although the stock price itself responded with a massive drop and volatility burst (Fig. 1). In 2014–2015, the confidence bands are slightly wider, reflecting the bubble and burst in Shanghai. In that period, the correlations sometimes go beyond 0.1 and outside the confidence bands. Hence, conditional on lag  $h = 1$ , there is a possibility that the Shanghai collapse had an adverse impact on the market efficiency.

The autocorrelation of Nikkei generally lies in  $[-0.1, 0.1]$  and they are insignificant in most windows, suggesting that the Nikkei market is weak form efficient. The correlation goes beyond 0.1 in only one out of 2910 windows, which is window #1772 (March 24, 2010–March 15, 2011). This is the first window that contains the tsunami shock.

For FTSE and SP500, we observe an interesting tendency that significantly negative autocorrelations sometimes emerge. For FTSE, the autocorrelation goes below  $-0.1$  for 954 windows out of 3004; a rejection occurs in 478 out of the 954 windows; the correlation even goes below  $-0.2$  for 106 windows, and a rejection occurs in 66 windows. Similar patterns arise in SP500. Recall that a rejection of the white noise hypothesis is evidence against weak form efficiency. Thus, U.K. and U.S. markets, which are broadly recognized as some of the most mature and liquid markets, are periodically inefficient, in particular during crisis periods.

Another implication from Figs. 5–7 is that the significantly negative correlations concentrate on periods of financial turmoil: the Iraq-War in 2003 for FTSE and SP500, and the subprime mortgage crisis in 2008 for SP500. This suggests that negative correlations may be indicative of

trading turmoil due to the rapid evolution of information. Evidence for negative correlations during crisis periods is not new. See, for example, Campbell et al. (1993) who find a negative relationship between trading volume and serial correlation: high volume days are associated with lower or negative correlations, due, they argue, to the presence of “non-informational” traders.

#### 4.2. White noise test

We perform the CvM white noise test for each rolling window. See Figs. 8–10 for bootstrapped p-values with window size  $n \in \{240, 480, 720\}$ , respectively. We use Shao's (2011) blockwise wild bootstrap with block size  $b_n = [c \times \sqrt{n}]$ . We draw  $c \sim U(0.5, 1.5)$  independently across  $M = 5000$  bootstrap samples and rolling windows. The shaded areas in the figures depict the nominal size  $p = 0.05$ . Also see Table 1, in which we report the ratio of rolling windows where the null hypothesis of white noise is rejected at the 5% level.

Our results suggest that the Chinese and Japanese stock markets are weak form efficient throughout the whole sample period. When window size is  $n = 240$  days, for example, a rejection happens in only 5.8% of all windows for Shanghai and 0.1% for Nikkei. Recall from Figs. 5–7 that Shanghai has positive correlations that are barely significant in 2014–2015. Those are only correlations at lag 1, and statistical significance declines when all lags are considered jointly. A rejection rarely happens under the larger window sizes  $n \in \{480, 720\}$  as well, suggesting the robustness of our finding.

FTSE and SP500 have more periods of inefficiency than Shanghai and Nikkei. When window size is  $n = 240$  days, for example, a rejection happens in as many as 14.3% of all windows for FTSE and 21.3% for SP500. The rejections occur continuously during the Iraq War and the subprime mortgage crisis (see Figs. 8–10). The former has a longer impact than the latter for FTSE, while the latter has a longer impact for SP500. This result is consistent with Figs. 5–7. It is also consistent

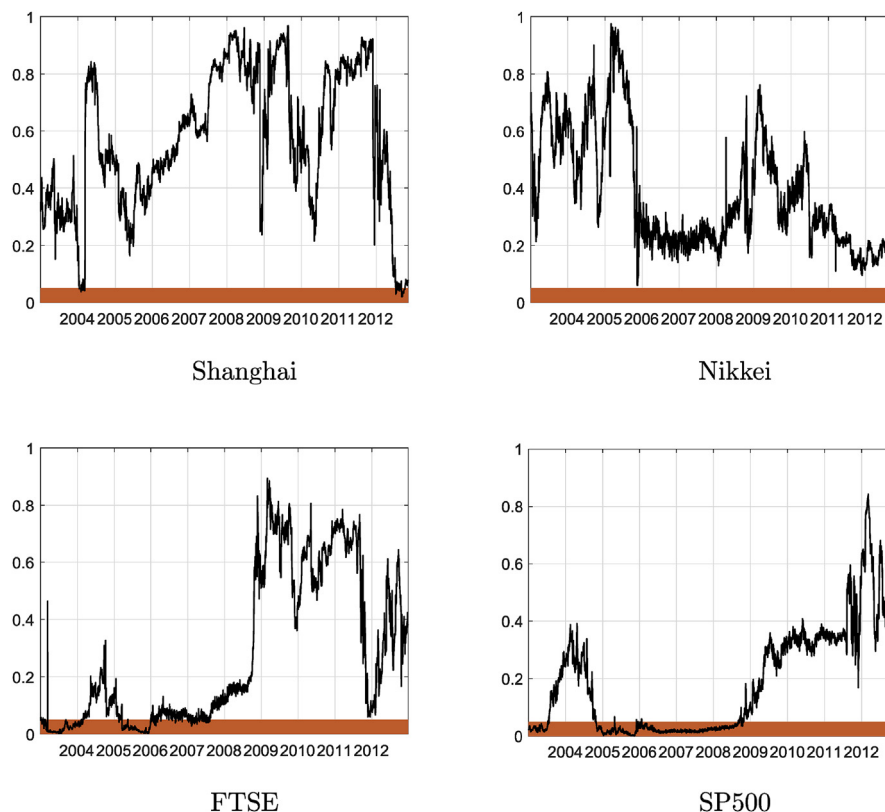


Fig. 10. P-Values of CvM Test with Window Size  $n = 720$  (Randomized Block Size).

Table 1

Rejection ratio of Cramér-von Mises test across rolling windows.

	Shanghai	Nikkei	FTSE	SP500
$n = 240$	0.058	0.001	0.143	0.213
$n = 480$	0.010	0.000	0.171	0.266
$n = 720$	0.023	0.000	0.227	0.443

with the notion that high volatility is associated with lower or negative correlations.

In Table 1, we observe that the rejection ratio of SP500 rises considerably as a window size increases: 21.3% for  $n = 240$ , 26.6% for  $n = 480$ , and 44.3% for  $n = 720$ . This result confirms our finding that the subprime mortgage crisis had a large adverse impact on the stock market efficiency in the U.S. To see that, focus on Fig. 10, where the window size is  $n = 720$  (roughly 3 years). P-values are below 0.05 for most windows containing at least a day of the year 2008, and the number of those windows accounts for approximately 39% of all windows. In general, the share of windows containing at least a day of a specific year rises as a window size increases. It is therefore not surprising that, given the large influence of the subprime mortgage crisis on the U.S. market, the rejection ratio of SP500 rises as a window size increases.

## 5. Conclusion

The vast majority of the literature on testing for weak form efficiency of stock markets imposes serial independence or a martingale difference assumption under the null hypothesis. Serial independence

rules out any form of conditional heteroskedasticity, and the mds property rules out higher level forms of dependence. It is thus of interest to perform white noise tests with little more than serial uncorrelatedness under the null hypothesis. A rejection of the white noise hypothesis is supposed to serve as a helpful signal of an arbitrage opportunity for investors, since it indicates the presence of non-zero autocorrelation at some lags.

We test for the white noise hypothesis of daily stock returns in Chinese, Japanese, U.K., and U.S. stock markets. Based on Shao (2011) and Hill and Motegi (2017), we perform the Cramér-von Mises test assisted by the blockwise wild bootstrap. In line with the adaptive market hypothesis, we run rolling window analysis to capture time-varying market efficiency. The present study is most likely the first to use of the blockwise wild bootstrap in a rolling window environment. The block structure inscribes an artificial periodicity in bootstrapped p-values or confidence bands across rolling windows, which we have successfully removed by randomizing the block size.

The degree of stock market efficiency differs noticeably across countries and sample periods, as asserted by the adaptive market hypothesis. For Shanghai and Nikkei, the white noise hypothesis is accepted in most windows, indicating that the Chinese and Japanese stock markets are weak form efficient throughout the sample period. The same goes for FTSE and SP500 as far as non-crisis periods are concerned. When FTSE and SP500 face greater uncertainty, namely the Iraq War and the subprime mortgage crisis, we tend to observe significantly negative autocorrelations. In those periods the white noise hypothesis is rejected, indicating that the U.K. and U.S. markets are inefficient. A negative correlation, in particular at low lags, signifies rapid changes in market trading, which is corroborated with high volatility due to non-informational traders.

## Appendix. Supplemental Analysis with Aggregated Data

In the main body of the paper, we analyze the *daily* stock prices. In this appendix, we analyze aggregated *weekly* and *monthly* stock prices to see how empirical results change. We compute weekly stock prices by picking a price on the last business day of each week. Similarly, we compute monthly stock prices by picking a price on the last business day of each month. Such an aggregation scheme is called *stock aggregation*, a natural way of aggregating stock variables like stock price indices.

In the main body of the paper, sample size is more than 3100 days for each market and the size of each rolling window is  $n \in \{240, 480, 720\}$  days (roughly 1, 2, and 3 years). Since aggregated data have much smaller sample sizes, we should reconsider the size of rolling windows. For weekly data, sample size is approximately 650 weeks and we set the window size to be  $n \in \{96, 144\}$  weeks (roughly 2 and 3 years). We do not consider  $n = 48$  weeks (roughly a year) since white noise tests would have poor performance in such a small sample. For monthly data, sample size is 153 months and we set the window size to be  $n = 96$  months (8 years). This is again a compromise to keep enough sample size for each rolling window.

As in the main body, we perform the Cramér-von Mises white noise test across rolling windows. We use the blockwise wild bootstrap with randomized block size. In Table A, we report the ratio of rolling windows where the null hypothesis of white noise is rejected at the 5% level (cf. Table 1). The null hypothesis is accepted for most windows, suggesting that each market is weak form efficient throughout the whole sample period. The market inefficiency observed for daily FTSE and SP500 in the crisis periods disappear at the weekly and monthly levels. It is an intuitively reasonable result since the larger time span should make it easier for investors to fully exploit arbitrage opportunities, resulting in the higher degree of market efficiency.

**Table A**  
Rejection ratio of Cramér-von Mises test over rolling windows (aggregated data).

	Weekly		Monthly
	$n = 96$	$n = 144$	$n = 96$
Shanghai	0.027	0.014	0.000
Nikkei	0.007	0.017	0.000
FTSE	0.004	0.004	0.000
SP500	0.066	0.049	0.000

## Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.econmod.2018.08.003>.

## References

- Anagnostidis, P., Varsakelis, C., Emmanouilides, C.J., 2016. Has the 2008 financial crisis affected stock market efficiency? The case of eurozone. *Physica A* 447, 116–128.
- Andrews, D.W.K., Ploberger, W., 1996. Testing for serial correlation against an ARMA(1,1) process. *J. Am. Stat. Assoc.* 91, 1331–1342.
- Billingsley, P., 1961. The Lindeberg-Lévy theorem for martingales. *Proc. Am. Math. Soc.* 12, 788–792.
- Campbell, J.Y., Grossman, S.J., Wang, J., 1993. Trading volume and serial correlation in stock returns. *Q. J. Econ.* 108, 905–939.
- Campbell, J.Y., Lo, A.W., MacKinlay, A.C., 1997. *The Econometrics of Financial Markets*. Princeton University Press.
- Charles, A., Darné, O., 2009. Variance-ratio tests of random walk: an overview. *J. Econ. Surv.* 23, 503–527.
- Charles, A., Darné, O., Kim, J.H., 2017. Adaptive markets hypothesis for Islamic stock indices: evidence from dow Jones size and sector-indices. *Int. Econ.* 151, 100–112.
- Delgado, M.A., Hidalgo, J., Velasco, C., 2005. Distribution free goodness-of-fit tests for linear processes. *Ann. Stat.* 33, 2568–2609.
- Delgado, M.A., Velasco, C., 2011. An asymptotically pivotal transform of the residuals sample autocorrelations with application to model checking. *J. Am. Stat. Assoc.* 106, 946–958.
- Deo, R.S., 2000. Spectral tests of the martingale hypothesis under conditional heteroscedasticity. *J. Econom.* 99, 291–315.
- Durlauf, S.N., 1991. Spectral based testing of the martingale hypothesis. *J. Econom.* 50, 355–376.
- Escanciano, J.C., Lobato, I.N., 2009. An automatic portmanteau test for serial correlation. *J. Econom.* 151, 140–149.
- Escanciano, J.C., Velasco, C., 2006. Generalized spectral tests for the martingale difference hypothesis. *J. Econom.* 134, 151–185.
- Fama, E.F., 1970. Efficient capital markets: a review of theory and empirical work. *J. Finance* 25, 383–417.
- Fama, E.F., French, K.R., 1988. Permanent and temporary components of stock prices. *J. Polit. Econ.* 96, 246–273.
- Guay, A., Guerre, E., Lazarova, S., 2013. Robust adaptive rate-optimal testing for the white noise hypothesis. *J. Econom.* 176, 134–145.
- Hill, J.B., Motegi, K., 2017. A Max-Correlation White Noise Test for Weakly Dependent Time Series. Working Paper. Department of Economics at the University of North Carolina at Chapel Hill.
- Hill, J.B., Motegi, K., 2018. “Supplemental Material for” Testing the White Noise Hypothesis of Stock Returns. Department of Economics at the University of North Carolina at Chapel Hill.
- Hong, Y., 1996. Consistent testing for serial correlation of unknown form. *Econometrica* 64, 837–864.
- Hong, Y., 1999. Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. *J. Am. Stat. Assoc.* 94, 1201–1220.
- Hong, Y., 2001. A test for volatility spillover with application to Exchange rates. *J. Econom.* 103, 183–224.
- Horowitz, J.L., Lobato, I.N., Nankervis, J.C., Savin, N.E., 2006. Bootstrapping the Box-Pierce Q test: a robust test of uncorrelatedness. *J. Econom.* 133, 841–862.
- Kim, J.H., Shamsuddin, A., 2008. Are Asian stock markets efficient? Evidence from new multiple variance ratio tests. *J. Empir. Finance* 15, 518–532.
- Kim, J.H., Shamsuddin, A., Lim, K.-P., 2011. Stock return predictability and the adaptive markets hypothesis: evidence from century-long U.S. Data. *J. Empir. Finance* 18, 868–879.
- Lahiri, S.N., 1999. Theoretical comparisons of block bootstrap methods. *Ann. Stat.* 27, 386–404.
- Lim, K.-P., Brooks, R., 2009. The evolution of stock market efficiency over time: a survey of the empirical literature. *J. Econ. Surv.* 25, 69–108.
- Lim, K.-P., Brooks, R.D., Kim, J.H., 2008. Financial crisis and stock market efficiency: empirical evidence from Asian countries. *Int. Rev. Financ. Anal.* 17, 571–591.
- Lim, K.-P., Luo, W., Kim, J.H., 2013. Are US stock index returns predictable? Evidence from automatic autocorrelation-based tests. *Appl. Econ.* 45, 953–962.
- Lo, A.W., 2004. The adaptive markets hypothesis: market efficiency from an evolutionary perspective. *J. Portfolio Manag.* 30, 15–29.
- Lo, A.W., 2005. Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis. *J. Invest. Consult.* 7, 21–44.
- Lo, A.W., MacKinlay, A.C., 1988. Stock market prices do not follow random walks: evidence from a simple specification test. *Rev. Financ. Stud.* 1, 41–66.
- Lobato, I.N., 2001. Testing that a dependent process is uncorrelated. *J. Am. Stat. Assoc.* 96, 1066–1076.
- Lobato, I.N., Nankervis, J.C., Savin, N.E., 2002. Testing for zero autocorrelation in the presence of statistical dependence. *Econom. Theor.* 18, 730–743.
- MacKinnon, J.G., 1996. Numerical distribution functions for unit root and cointegration tests. *J. Appl. Econom.* 11, 601–618.
- Mantalos, P., Shukur, G., 2008. Bootstrap methods for autocorrelation test with uncorrelated but not independent errors. *Econ. Modell.* 25, 1040–1050.
- Mitra, S.K., Chattopadhyay, M., Charan, P., Bawa, J., 2017. Identifying periods of market inefficiency for return predictability. *Appl. Econ. Lett.* 24, 668–671.
- Nankervis, J.C., Savin, N.E., 2010. Testing for serial correlation: generalized Andrews-Ploberger tests. *J. Bus. Econ. Stat.* 28, 246–255.

- Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance Matrix estimation. *Rev. Econ. Stud.* 61, 631–653.
- Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. *Biometrika* 75, 335–346.
- Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. *J. Am. Stat. Assoc.* 89, 1303–1313.
- Romano, J.P., Thoms, L.A., 1996. Inference for autocorrelations under weak assumptions. *J. Am. Stat. Assoc.* 91, 590–600.
- Shao, X., 2010. The dependent wild bootstrap. *J. Am. Stat. Assoc.* 105, 218–235.
- Shao, X., 2011. A bootstrap-assisted spectral test of white noise under unknown dependence. *J. Econom.* 162, 213–224.
- Urquhart, A., McGroarty, F., 2016. Are stock markets really efficient? Evidence of the adaptive market hypothesis. *Int. Rev. Financ. Anal.* 47, 39–49.
- Verheyden, T., De Moor, L., Van den Bossche, F., 2015. Towards a new framework on efficient markets. *Res. Int. Bus. Finance* 34, 294–308.
- Wright, J.H., 2000. Alternative variance-ratio tests using ranks and signs. *J. Bus. Econ. Stat.* 18, 1–9.
- Yen, G., Lee, C., 2008. Efficient market hypothesis (EMH): past, present and future. *Rev. Pac. Basin Financ. Mark. Policies* 11, 305–329.
- Zhang, X., 2016. White noise testing and model diagnostic checking for functional time series. *J. Econom.* 194, 76–95.
- Zhu, K., Li, W.K., 2015. A bootstrapped spectral test for adequacy in weak ARMA models. *J. Econom.* 187, 113–130.