

Tan, B. and Yilmaz, K., “[Markov Chain Test for Time Dependence and Homogeneity: An Analytical and Empirical Evaluation](#),” *European Journal of Operational Research*, Vol. 137, No. 3, pp. 524-543, 2002.

DOI: 10.1016/S0377-2217(01)00081-9

©2020. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
<http://creativecommons.org/licenses/by-nc-nd/4.0/>



ELSEVIER

European Journal of Operational Research 137 (2002) 524–543

EUROPEAN
JOURNAL
OF OPERATIONAL
RESEARCH

www.elsevier.com/locate/dsw

Stochastics and Statistics

Markov chain test for time dependence and homogeneity: An analytical and empirical evaluation

Barış Tan ^{*,1}, Kamil Yılmaz

Koç University, Rumeli Feneri Yolu, Sarıyer İstanbul 80910, Turkey

Received 7 October 1999; accepted 8 February 2001

Abstract

This paper evaluates the small and large sample properties of Markov chain time-dependence and time-homogeneity tests. First, we present the Markov chain methodology to investigate various statistical properties of time series. Considering an auto-regressive time series and its associated Markov chain representation, we derive analytical measures of the statistical power of the Markov chain time-dependence and time-homogeneity tests. We later use Monte Carlo simulations to examine the small-sample properties of these tests. It is found that although Markov chain time-dependence test has desirable size and power properties, time-homogeneity test does not perform well in statistical size and power calculations. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Time series; Markov chain; Monte Carlo simulation; Efficient market hypothesis

1. Introduction

Over the last forty years, Markov chain based tests have been frequently used in the analysis of statistical properties of various economic time series. Market efficiency literature is the area that made the most use of Markov chain based tests. Among others, Niederhoffer and Osborne (1966), Dryden (1969), Fielitz and Bhargava (1973), McQuenn and Thorley, 1991) applied Markov

chain based tests on stock markets, whereas Gregory and Sampson (1987) tested market efficiency in the foreign exchange markets.

Markov chain based tests require transforming a continuous state space process associated with a given time series into a discrete state space sequence. If only the time-dependency and homogeneity properties of a time series are of interest, observations can be aggregated into a finite number of states. Then, this state space representation can be used to investigate these properties.

Similar to all statistical testing procedures, there are pros and cons of using Markov-chain based tests. First, there is a trade-off between gaining more information about the structural properties of the time series and losing the details when a continuous state space of a given time

^{*} Corresponding author. Tel.: +90-212-229-3006; fax: +90-212-229-0680.

E-mail address: btan@ku.edu.tr (B. Tan).

¹ Present address: Harvard University, Maxwell Dworkin, 317, 33 Oxford Street, Cambridge, MA 02138, USA; Tel.: +1-617-495-9875; fax: +1-617-495-9837.

series is aggregated into a discrete state one. Second, Markov chain based tests do not require an extensive set of assumptions about the distribution, homoscedasticity, serial correlation, etc. However, Markov chain time-dependence test requires that the Markov chain associated with the time series under investigation is time-homogeneous. That is, it is required that the state transition probabilities do not change over time. Although time-homogeneity can be tested as well, testing time-dependency together with time-homogeneity introduces some challenges for Markov chain based tests.

Most of the previous studies that used Markov chains to test market efficiency ignored the interdependence between the time-dependence (relevant for market efficiency) and time-homogeneity properties of the Markov chain representation of stock returns. Having focused on market efficiency, these studies tested only time-dependence, while assuming time-homogeneity without testing, or tested time-homogeneity by assuming a given order of time-dependence. However, overlooking this interdependence may lead to distortions in the test results. For example, we believe that assuming time-homogeneity of stock returns over decades without testing, and then commenting on market efficiency based on the Markov chain test results is not methodologically appropriate.

This shortcoming in the literature motivated us to evaluate the performance of the Markov chain time-dependence and time-homogeneity tests. Towards this end, we first present the Markov chain testing methodology. Then we report the results of the analytical and empirical evaluation of the tests. The remainder of the study is organized as follows. Previous studies on Markov chain based statistical tests and applications to stock markets are reviewed in Section 2. A detailed framework to use this methodology to test various properties of time series is presented in Section 3. In Section 4, we report the analytical evaluation of the Markov chain as a test of time-dependence and homogeneity. Monte Carlo simulation results on the size and power of Markov chains tests are presented in Section 5. Finally, Section 6 concludes the paper.

2. Review of past work

As Andrew Lo (1997) states, efficient market hypothesis (EMH) is perhaps "... one of the most controversial and well-studied propositions in all the social sciences." The research on the EMH, dates back to 1960s, yet empirical work on stock price behavior has not produced a consensus on whether or not stock prices follow random walk, an implication of the EMH. Most of the earlier research (Osborne, 1959; Cootner, 1962; Fama, 1965; Fama and Blume, 1966) failed to reject random walk hypothesis (RWH) for short-term returns, using daily or weekly stock price data. Yet, other studies rejected RWH for daily returns and intra-day price changes in the New York stock exchange (Niederhoffer and Osborne, 1966; Fielitz and Bhargava, 1973; Fielitz, 1975).

The interest in the subject flared in the second half of 1980s, with the focus shifting from high frequency to low frequency stock returns. Using regression and variance ratio (VR)-tests, respectively, Fama and French (1988) and Poterba and Summers (1988) found support for mean-reversion in monthly US stock prices at 3–5-year return horizons. These findings and the testing methodology were later challenged by Kim et al. (1991) and Richardson (1993). They showed that evidence for mean-reversion in stock prices became weaker once the test statistic is adjusted to reflect the statistical properties of the underlying stock return data. While the debate on long-term behavior of stock return was taking place, Lo and MacKinlay (1988) showed that unlike long-term stock returns weekly US stock returns displayed positive serial correlation for short return horizons.

One common problem with these tests is that in most cases rejection occurs at relatively high marginal significance levels and they have generally low power against alternative hypotheses. Another is the underlying assumption of linear dependence between current and past stock returns. Considering this shortcoming, many authors used the alternative discrete Markov chain test procedure to test for random walk behavior. With the possibility of order of dependence greater than one, Markov chain test allows one to

consider non-linear time-dependence of stock returns as an alternative to the random walk. Niederhoffer and Osborne (1966), Fielitz and Bhargava (1973), Fielitz (1975) and McQuenn and Thorley (1991) applied Markov chain test to individual stock prices as well as market indices at various frequencies for different time periods.

The studies based on Markov chains generated more support against, than in favor of, the RWH. The majority of these studies used high frequency data (weekly and daily returns as well as intra-day ticker price changes) and made use of large sample properties. Using daily and weekly returns for 200 individual stocks between 1963 and 1968 and a three-state Markov chains, Fielitz and Bhargava (1973) tested for the stationarity (time-homogeneity) as well as the order of dependence of the Markov chain. They obtained several results. First, they showed that these 200 stocks could not be treated as generated by a single vector Markov chain. More importantly, they showed that, the behavior of individual stock returns were not governed by a stationary Markov chain process. That is to say, the transition probability matrix for these stock returns is not stationary throughout the 1963–1968 period. Despite the rejection of stationarity, however, their order of dependence tests provide evidence for short-term (first- or higher-order) dependence in daily returns of individual stocks, but no dependence in weekly Monday closing returns.

Niederhoffer and Osborne (1966) applied Markov chain test to high frequency, ticker price changes from one transaction to another of seven stocks included in the Dow Jones industrial average index for a 22-day trading period in October 1964. They divided the ticker price changes into seven states, where the distance between each state is 1/8 of a dollar. For these seven stocks, transaction price changes turned out to have a stationary transition probability matrix. Using a first-order Markov chain they rejected random walk and consistently showed that the reversal of a price change was more likely to take place than a continuation of the change.

Finally, McQuenn and Thorley (1991) recently applied Markov chain test to annual real and excess returns of the equally weighted and value-

weighted portfolios of all stocks in the New York stock exchange between 1947 and 1987. Based on earlier research, they adopted a two-state (up–down) Markov chain. They implicitly assumed that the transition probability matrix was stationary throughout the period of analysis, and tested random walk against second-order dependence. In this setting, they rejected the RWH and obtained evidence in favor of mean-reversion in long-horizon stock returns for the postwar period. Their results are stronger than earlier findings of mean-reversion in annual stock returns by Poterba and Summers (1988) and Fama and French (1988) for the pre-war period. With 40 observations, McQueen and Thorley showed that their results are not due to small sample bias.

The studies based on Markov chain tests altogether provided ample evidence against random walk. However, none of the papers considered the Markov chain methodology in its entirety. They applied Markov chain tests without taking full account of the properties of the stock returns data they were using. Unlike other tests of random walk behavior, Markov chains can be used to detect possible changes in the behavioral pattern of the stock returns over time. A significant break in the time series behavior of stock returns would generate rejection of the stationarity of the transition probability matrix (a.k.a. time-homogeneity of Markov chains). For this reason, testing for the stationarity of the transition probability matrix plays an essential role for the validity of Markov chain based random walk tests. Even if Markov chain test rejects the RWH against the alternative of a first- or higher-order return dependence, that relationship cannot be claimed stationary through time without proper testing. If the transition probability matrix is not stationary through time, then the Markov chain will have no predictive power. In such a case the correct approach would be to apply Markov chains test to sub-sample periods and test whether the transition probability matrices for these sub-sample periods are stationary or not. Tests for the order of dependence and hence for random walk would be appropriate for those sub-sample periods where the stationarity of the transition probability matrix cannot be rejected.

Even though, they were considering a period of 40 years through which the NYSE went through substantial changes, McQuenn and Thorley (1991) did not test for the stationarity of transition probability matrix. In their tests of random walk vs second-order dependence, they maintained the time-homogeneity assumption. They did not consider testing the RWH against the alternative of first- or higher-order dependence. Moreover, they used a two-state Markov chain without checking if their results were robust to the choice of number of states.

Fielitz and Bhargava (1973) and Fielitz (1975) were aware of the importance of time-homogeneity property for the prediction capacity of Markov chains. Their tests showed that the transition probability matrices were not stationary over time, yet they went ahead, tested for and rejected random walk in favor of first- or higher-order dependence.

Another study that addressed the importance of the time-homogeneity property of Markov chains is Gregory and Sampson (1987). They used Markov chain time-dependence test to investigate the independence of forecast errors in monthly forward foreign exchange rates for British, Canadian, French, German, Italian and Japanese currencies vis-a-vis US dollar between 1974 and 1981. The Markov chains test rejected the temporal independence of forward foreign exchange forecast errors for all currencies, except for the Canadian Dollar. Unlike other studies cited above, Gregory and Sampson (1987) first tested for the time-homogeneity property of the Markov chain before reaching any conclusion about the time-dependence of forecast errors.

3. Markov chain based testing methodology

In this study, we follow a Markov chain based methodology with the purpose of investigating the time-dependency and homogeneity properties of a given time series. This methodology is a non-parametric testing procedure that estimates the probability distribution directly. Furthermore, it does not require the distributional assumptions that are required for parametric tests.

Formally, we view a given time series, $y(t)$, as a discrete parameter, continuous state space stochastic process $\{y(t), t = 0, 1, 2, \dots\}$. In order to utilize the Markov chain tests, we aggregate the continuous state space of the time series into a discrete state space with a finite number of states. That is, the process $\{y(t), t = 0, 1, 2, \dots\}$ is mapped into a discrete parameter, discrete state space stochastic process denoted as $\{X_t, t = 0, 1, 2, \dots\}$ on the state space S of size n_s .

The aggregation of the state space and the definition of the state space depend on the statistical properties of the time series under investigation. Many studies that analyze the statistical properties of stock-price movements focus on the *direction* of the movements, e.g., whether a stock price goes up or down (with respect to, say, the average return) at a given time and how these movements are related to the stock price's previous movements and time (see, among others, Fielitz and Bhargava, 1973; McQuenn and Thorley, 1991). It is possible to examine these properties by only looking at the *directions* rather than the stock *prices*. In this case, the continuous state space of stock prices are mapped into a discrete state space $S = \{U, D\}$, where U corresponds to an upward movement of $y(t)$ at time t , D corresponds to a downward movement of $y(t)$ with respect to its average during the full period $[0, T]$, $\bar{y} = (1/(T+1)) \sum_{t=0}^T y(t)$, i.e.,

$$X_t = \begin{cases} U & \text{if } y(t) \geq \bar{y}, \\ D & \text{if } y(t) < \bar{y}. \end{cases} \quad (1)$$

In addition to the direction of movement, if the *magnitude* is of interest, then the number of states in the state space can be increased to gain more information. In the simplest case, including one additional state that corresponds to no upward or downward movement of $y(t)$ at time t yields some information about the magnitude of change. That is, a three-state stochastic process on state space $X_t \in S$, $S = \{U, N, D\}$, where U corresponds to an upward movement, N corresponds to no upward or downward movement (stay the same) and D corresponds to a downward movement of $y(t)$, can be used. The intervals that correspond to the states U , N , and D can be determined by dividing the

frequency distribution of $y(t)$ into three equiprobable regions (Dryden, 1969).

Alternatively one can use more states in S to include more information on $y(t)$ in X_t (for example, Niederhoffer and Osborne, 1966). However, including more states increases the number of parameters to be estimated and thus reduces the testing power when the number of observations is limited.

Fig. 1 depicts sample realizations of the stochastic processes $\{y(t), t = 0, 1, 2, \dots, T\}$, $\{X_t, t = 0, 1, 2, \dots, T\}$, $X_t \in \{U, D\}$ and $\{X'_t, t = 0, 1, 2, \dots, T\}$, $X'_t \in \{U, N, D\}$ for a specific stock market index.

3.1. Testing time-dependence

If a given sequence is an independent process or random walk, then the movements at any given

time are independent of each other. Thus testing the random walk of the stock market hypothesis is equivalent to testing the independence of the stochastic process. By definition, a given sequence $\{X_t, t = 0, 1, 2, \dots\}$ is an independent process if for all $t, t = 0, 1, 2, \dots$, the probability law of the process is given by

$$\begin{aligned} P[X_t = j, X_{t-1} = i_1, \dots, X_{t-u} = i_u, \dots, X_0 = i_t] \\ = P[X_t = j] \cdot P[X_{t-1} = i_1] \cdots P[X_0 = i_t] \end{aligned} \quad (2)$$

or equivalently

$$P[X_t = j | X_{t-1} = i_1, \dots, X_0 = i_t] = P[X_t = j]. \quad (3)$$

If the above condition does not hold, then determining the degree of dependency, that is, whether a given movement depends on the last movement, last two movements, etc., is of interest. If $\{X_t, t = 0, 1, 2, \dots\}$ is a first-order Markov chain, or simply a Markov chain, then

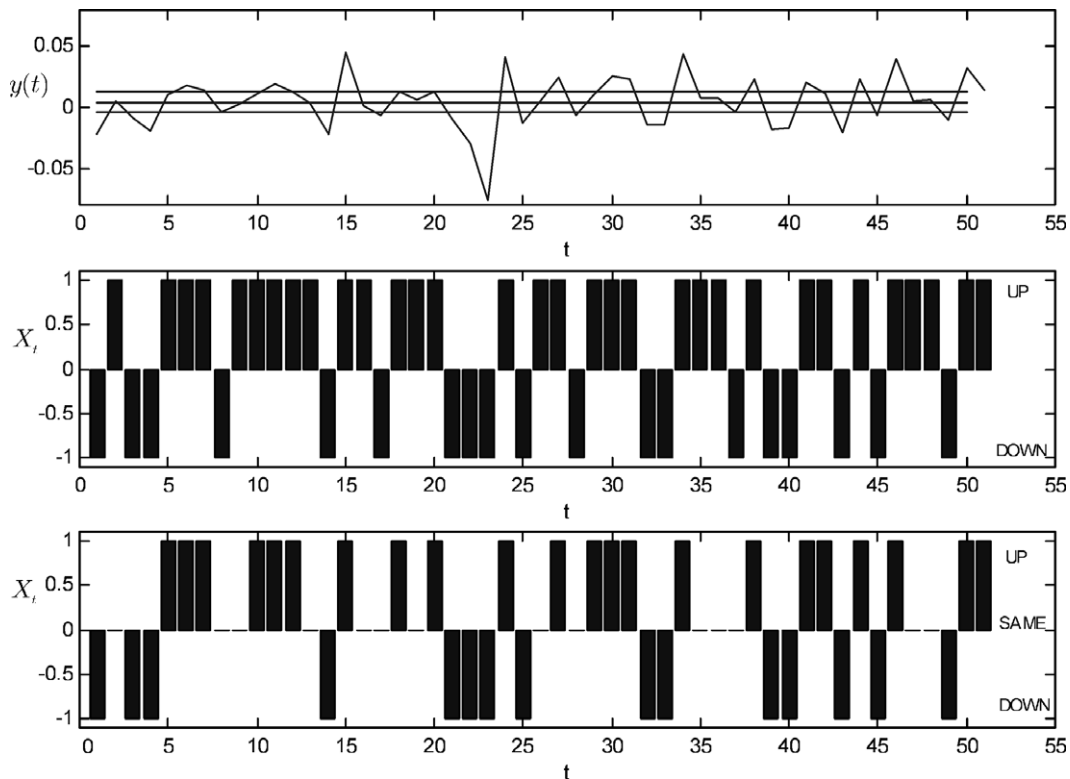


Fig. 1. Sample realizations of processes $\{y(t), t = 0, 1, 2, \dots, T\}$, $\{X_t, t = 0, 1, 2, \dots, T\}$ for a specific stock market index.

$$\begin{aligned}
 P[X_t = j | X_{t-1} = i_1, \dots, X_0 = i_t] \\
 = P[X_t = j | X_{t-1} = i_1], \quad t = 1, 2, \dots
 \end{aligned} \quad (4)$$

Similarly, if $\{X_t, t = 0, 1, 2, \dots\}$ is a Markov chain of order u , then

$$\begin{aligned}
 P[X_t = j | X_{t-1} = i_1, \dots, X_{t-u} = i_u, \dots, X_0 = i_t] \\
 = P[X_t = j | X_{t-1} = i_1, \dots, X_{t-u} = i_u], \\
 t = u, u+1, u+2, \dots
 \end{aligned} \quad (5)$$

A time-homogeneous Markov chain of order u is completely characterized with its state transition matrix $P = \{p_{i,j}\}$, where

$$\begin{aligned}
 p_{i,j} = P[X_t = j | X_{t-1} = i_1, \dots, X_{t-u} = i_u], \\
 i = (i_1, \dots, i_u) \in S^u, \quad j \in S, \quad t = 0, 1, 2, \dots
 \end{aligned} \quad (6)$$

In this representation state i includes more than one state if the order of time-dependency is greater than one. For example, for a second-order Markov chain defined on the state space $\{U, D\}$, $i \in \{UU, UD, DU, DD\}$ and $j \in \{U, D\}$.

We now briefly present the procedure used to estimate the state transition probabilities and to test the order of the Markov chain for the completeness of the methodology. For more detailed information about these procedures, the reader is referred to Anderson and Goodman (1957).

Once it is assured that state transition probabilities do not change with time, i.e., the Markov chain is time-homogenous, over a given period, and the order of the Markov chain is set then these probabilities can be estimated directly from the observed transitions. The maximum likelihood estimates of the state transition probabilities are obtained by maximizing the maximum likelihood function subject to the constraint that $\sum_j p_{i,j} = 1$, $i \in S^u$, $j \in S$, which yields

$$\hat{p}_{i,j} = \frac{n_{i,j}}{n_i}, \quad i \in S^u, \quad j \in S, \quad (7)$$

where $n_{i,j}$ is the total number of observed transitions from state $i \in S^u$ to $j \in S$, and n_i is the total number of transitions from state i during a given time period. For small samples, the above equation can be adjusted to incorporate the effect of the

initial distribution (e.g., Neftci, 1984; Sherlaw-Johnson et al., 1995).

In the testing procedure, we test the null hypothesis that the Markov chain is of order u vs the alternative that it is of order v , $u < v$. Let $P = \{p_{i,j}\}$ and $Q = \{q_{i,j}\}$ denote the time-homogeneous state transition matrices of Markov chains of orders u and v , respectively. Let $n_{i,j}$ be the total number of observed transitions from state $i \in S^v$ to $j \in S$. An asymptotically equivalent test statistic for the likelihood ratio test statistic, is given by

$$\begin{aligned}
 -2 \ln(A) = 2 \sum_{i,j} n_{i,j} [\ln(q_{i,j}) - \ln(\tilde{q}_{i,j})], \\
 i \in S^v, \quad j \in S,
 \end{aligned} \quad (8)$$

where

$$\tilde{Q} = \{\tilde{q}_{i,j}\} = \underbrace{[P^T, P^T, \dots, P^T]^T}_{2^{v-u}}.$$

This test statistic has a χ^2 asymptotic distribution with $(n_s^v - n_s^u)(n_s - 1)$ degrees of freedom.

Testing the independence of transition probabilities, i.e., random walk, vs the alternative hypothesis that the stochastic process is a Markov chain of order v is equivalent to testing order 0 vs order v by using the above given procedure.

We determine the order of the Markov chain for a given sequence by conducting a series of tests starting with order 0 vs order 1 and continuing, in the case of the rejection of the previous one, with order u vs $u+1$, $u = 1, 2, \dots$. That is, the order is first set to zero and the state transition probabilities for order 0 and order 1 are estimated. Then the hypothesis that the Markov chain is order 0 vs alternative hypothesis that the Markov chain is order 1 is tested. If this hypothesis cannot be rejected, it is concluded that the order is 0 and the time-homogeneity test is conducted next. However, if the hypothesis is rejected, then the current order is increased by one and the same test is applied until either the hypothesis cannot be rejected or a prespecified limit is reached. As the order increases, the number of parameters to be estimated increases exponentially and the null hypothesis is rejected due to decreased power.

Therefore, it is necessary to set a limit to the maximum order to be tested.

Note that, in order to implement the testing methodology outlined above, it is necessary to assure that the transition probabilities do not change with time, i.e., the underlying Markov chain is time-homogeneous. However, as the time period under investigation gets longer, assuming time-homogeneity will not be a plausible assumption. We discuss testing time-homogeneity of a Markov chain in the following section.

3.2. Testing time-homogeneity

Given that only one replication of a time series is available, in order to test time-homogeneity, we divide $\{X_t, t = 0, 1, 2, \dots, T\}$ into K different equal length sub-intervals and test whether the transition probabilities estimated for each period are statistically different from the transition probabilities estimated for the full period. Let the state transition probability of an u th order Markov chain corresponding to period k $k = 1, 2, \dots, K$ is given by

$$p_{i,j}(k) = P[X_t = j | X_{t-1} = i_1, \dots, X_{t-u} = i_u],$$

$$i = (i_1, \dots, i_u) \in S^u, \quad j \in S, \quad t \in [(k-1)\Delta, k\Delta],$$
(9)

where $\Delta = \lfloor (T+1)/K \rfloor$. We would like to test the hypothesis that the state transition matrices in each period, $P(k)$ are not statistically different from the transition matrix for the entire period $P = \{p_{i,j}\}$ vs the alternative hypothesis that they are different. Then an asymptotically equivalent test statistic for the likelihood ratio test statistic is given by

$$-2 \ln(A) = 2 \sum_k \sum_{i,j} n_{i,j}(k) [\ln(p_{i,j}(k)) - \ln(p_{i,j})],$$

$$i \in S^u, \quad j \in S, \quad k = 1, 2, \dots, K.$$
(10)

This test statistic has a χ^2 asymptotic distribution with $(K-1)n_s(n_s-1)$ degrees of freedom.

If this test is rejected, the process cannot be assumed to be time-homogeneous. Thus, no other tests regarding the order or the random walk

behavior can be performed by using a single transition matrix estimated from the observed transitions. Following the rejection of time-homogeneity for the entire period, the same test can be applied for each sub-period by dividing each sub-period further into smaller intervals. Note that if the number of observed transitions in each sub-period is low, than it is more likely to reject the null hypothesis against the alternative hypothesis mainly because of insufficient number of observations.

In order to test time-dependence and time-homogeneity together, we first determine the order by assuming time-homogeneity, and then test the hypothesis of time-homogeneity with the order determined in the previous step. Based on the result of the time-homogeneity test, it is possible to reach three different conclusions.

In the first case, the test estimates an order and accepts the null of time-homogeneity. That is, we conclude that the underlying Markov chain is time-homogenous and furthermore it has the specified order.

A second possibility is the case where time-homogeneity is rejected. This implies that the order of dependence cannot be estimated with Markov chains for the sample considered. We can proceed by dividing the period into sub-intervals and conducting the same procedure again. As a result, we can conclude, for example, a part of the period is time-homogeneous, i.e., the state transition probabilities do not change with time in this part, and the order dependency is one, while the order of the dependency for the other part of the full period is two.

Finally, the test may reject all orders of dependence equal to or less than the maximum order prespecified by the researcher. In what follows, we consider this case as *inconclusive*, because we cannot conclude about the order, and therefore cannot proceed with the time-homogeneity test. Fig. 2 depicts the flow chart of the Markov chain based testing methodology.

Given the Markov-chain-based joint time-dependency and homogeneity testing methodology, we next discuss some of its properties and its performance to analyze statistical properties of time series.

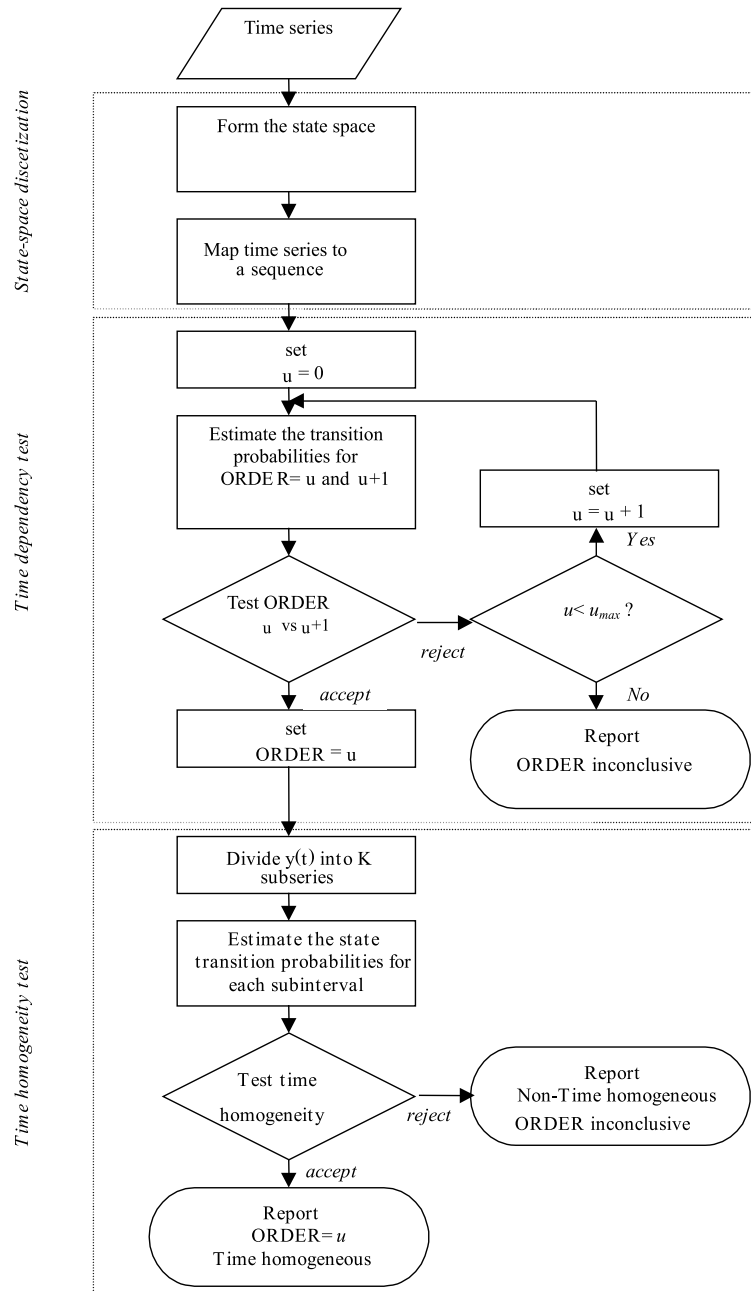


Fig. 2. Flowchart of combined time-dependence and homogeneity test.

4. An analytic evaluation of Markov chain tests of time series

Let us consider a simple first-order auto-regressive process, AR(1), $\{y(t), t \geq 0\}$ defined as

$$y(t) = \alpha y(t-1) + \varepsilon(t), \quad (11)$$

where $y(0) = 0$ and $\varepsilon(t)$ is a white noise process with mean 0 and variance σ^2 . Let us also consider

its first-order Markov chain representation $\{X_t, t = 0, 1, 2, \dots, T\}$, $X_t \in \{U, D\}$, where

$$X_t = \begin{cases} U & \text{if } y \geq (t)\bar{y}, \\ D & \text{if } y(t) < \bar{y} \end{cases} \quad (12)$$

and $\bar{y} = E[y(t)]$. In words, we observe the upward and downward movements of the process $y(t)$ and construct a Markov chain that describes the probabilities of one step transitions from an upward or downward movement to upward or downward movement. The Markov chain is completely characterized by its transition matrix $P = \{p_{ij}\}$, where

$$p_{ij} = P[X_{t+1} = j | X_t = i], \quad j \in \{U, D\}, \quad t = 0, 1, 2, \dots \quad (13)$$

These transition probabilities can be estimated from an observed realization of the time series.

Let us first assume that the time series can be observed for an infinitely long period. By the law of large numbers, the transition probabilities of the Markov chain will approach the true conditional probabilities of the process. That is, with probability one

$$p_{UU} = P[y(t+1) \geq \bar{y} | y(t) \geq \bar{y}]. \quad (14)$$

Furthermore, since the distribution of $y(t)$ is symmetric, the Markov chain representation of an AR(1) process is also symmetric: $p_{DD} = p_{UU}$ and $p_{UD} = p_{DU} = 1 - p_{UU}$. Therefore, determining a single probability measure, p_{UU} , describes the Markov chain completely. Under certain assumptions, it is possible to determine this probability analytically. First note that, $\bar{y} = E[y(t)] = 0$ and thus $P[y(t) > \bar{y}] = 1/2 \forall t$. Then

$$\begin{aligned} p_{UU} &= P[y(t) > \bar{y} | y(t-1) > \bar{y}] \\ &= \frac{P[y(t) > \bar{y}, y(t-1) > \bar{y}]}{P[y(t-1) > \bar{y}]}. \end{aligned} \quad (15)$$

The probability given in the numerator of the above equation can be calculated from the joint distribution of $y(t)$ and $y(t-1)$. It can be shown that the joint distribution of $y(t)$ and $y(t-1)$ is a bivariate normal with density function

$$f_{y(t), y(t-1)}(z, w) = \frac{\sqrt{1 - \alpha^2}}{2\pi\sigma^2} e^{-(1/2\sigma^2)(z^2 - 2\alpha zw + w^2)}. \quad (16)$$

Therefore, p_{UU} can be determined as a function of α as

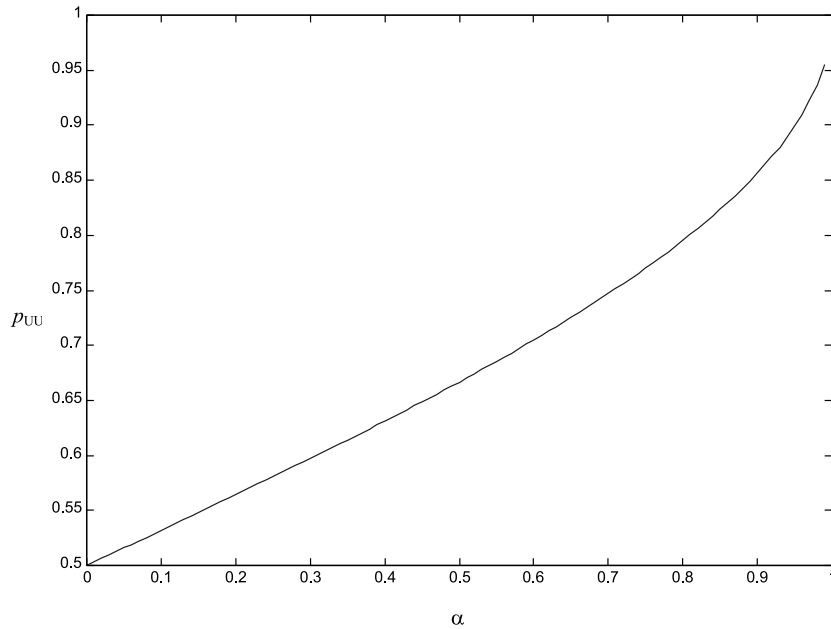
$$\begin{aligned} p_{UU}(\alpha) &= \frac{\sqrt{1 - \alpha^2}}{\pi\sigma^2} \int_0^\infty \int_0^\infty \\ &\times \exp\left(-\frac{1}{2\sigma^2}(z^2 - 2\alpha zw + w^2)\right) dz dw. \end{aligned} \quad (17)$$

Since the probability function of bivariate normal is not available in closed form, the transition probability is determined numerically. Fig. 3 plots the values of p_{UU} corresponding to different values of α .

Fig. 3 depicts that for a stock price series that follow random walk process ($\alpha = 0$) it is equally likely for the stock price to go up or down next period regardless of the current state. Similarly as α approaches one, the transition probability p_{UU} also approaches one. In other words, the process becomes more persistent. Furthermore, the figure suggests that the relationship between p_{UU} and α is one-to-one, there is a unique p_{UU} for each α value. Then, the parameter of AR(1) process can be estimated from the estimated transition probability of its Markov chain representation and vice versa.

Using Monte Carlo experiments, we investigated whether the analytical one-to-one correspondence between p_{UU} and α is valid for finite samples empirically. These experiments produced a similar relationship between the two for different sample sizes. Table 1 summarizes these results. The empirical p_{UU} values converge to the analytical p_{UU} values as the sample size increases from 100 to 250, 500 and 1000. The validation of the analytical results by Monte Carlo experiments strengthens our conclusion that it is possible to estimate an AR(1) process using Markov chains.

The power of the Markov chain test to detect random walk can be calculated using the analytical approach outlined above. Let us consider an AR(1) process of length N . The transition probabilities corresponding to this series are calculated by using the procedure explained above. The random walk test is equivalent to testing the order 0 vs order 1 for this Markov chain. The matrices

Fig. 3. p_{UU} vs α .Table 1
 p_{UU} vs a theoretical and simulation results

α	Theoretical $N = \infty$	Simulation			
		$N = 100$	$N = 250$	$N = 500$	$N = 1000$
0.1	0.5319	0.5256 (0.5220, 0.5292) ^a	0.5300 (0.5277, 0.5323)	0.5312 (0.5296, 0.5328)	0.5317 (0.5306, 0.5328)
0.2	0.5641	0.5598 (0.5563, 0.5633)	0.5628 (0.5607, 0.5650)	0.5629 (0.5613, 0.5645)	0.5643 (0.5632, 0.5654)
0.3	0.5970	0.5912 (0.5877, 0.5947)	0.5954 (0.5933, 0.5975)	0.5956 (0.5940, 0.5972)	0.5969 (0.5958, 0.5979)
0.4	0.6310	0.6237 (0.6203, 0.6272)	0.6277 (0.6255, 0.6299)	0.6305 (0.6289, 0.6320)	0.6307 (0.6296, 0.6318)
0.5	0.6667	0.6590 (0.6556, 0.6624)	0.6637 (0.6615, 0.6659)	0.6647 (0.6631, 0.6662)	0.6661 (0.6650, 0.6671)
0.6	0.7048	0.6989 (0.6956, 0.7022)	0.7012 (0.6991, 0.7033)	0.7024 (0.7009, 0.7039)	0.7044 (0.7033, 0.7055)
0.7	0.7468	0.7399 (0.7365, 0.7433)	0.7416 (0.7394, 0.7437)	0.7443 (0.7428, 0.7459)	0.7463 (0.7452, 0.7473)
0.8	0.7952	0.7845 (0.7811, 0.7879)	0.7902 (0.7881, 0.7923)	0.7929 (0.7915, 0.7944)	0.7940 (0.7930, 0.7951)
0.9	0.8564	0.8401 (0.8368, 0.8434)	0.8498 (0.8477, 0.8520)	0.8518 (0.8503, 0.8533)	0.8542 (0.8532, 0.8553)

^a 95% Confidence interval of p_{UU} obtained from 1000 replications of Monte Carlo simulation.

$Q = \{q_{i,j}\}$ and $\tilde{Q} = \{\tilde{q}_{i,j}\}$ that are used in the test statistic given in Eq. (8) are given as

$$Q = \begin{bmatrix} p_{UU}(\alpha) & 1 - p_{UU}(\alpha) \\ 1 - p_{UU}(\alpha) & p_{UU}(\alpha) \end{bmatrix},$$

$$\tilde{Q} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix},$$

respectively. Let us approximate the number of observations for each transition proportional to

these transition probabilities. Note that since $P[y(t) > \bar{y}] = 1/2$, we expect to see half of the observations from state U and the other half from state D . Then, $n_{i,j} = (N/2)p_{i,j}$. Finally, Eq. (8) can be written as a function of α as

$$\begin{aligned} -2 \ln(\mathcal{A}) = & N[p_{UU}(\alpha) \ln(2p_{UU}(\alpha)) \\ & + (1 - p_{UU}(\alpha)) \ln(2 - 2p_{UU}(\alpha))], \end{aligned} \quad (18)$$

which has a χ^2 distribution with one degree of freedom. By using Eq. (17), the p -value associated with the above test statistic can be determined numerically. Fig. 4 depicts the p -value of the random walk test for different values of α as the sample size increases. The figure indicates that the Markov chain test can distinguish an AR(1) process with $\alpha = 0.1$ from a random walk only if there are around 2000 samples at 5% significance level. However, as α increases, the test can reject the RWH by using much less number of observations.

One necessary condition for the validity of the Markov chain time-dependence test is the time-homogeneity of the underlying Markov chain, which itself can be tested as well. As described above, the time-homogeneity test for Markov chain can be viewed as a structural break test for the underlying time series. The power of the Markov chain test to detect the structural break can be calculated using the analytical approach outlined above. Let us consider a time series of length N . Let us further assume that the first half of this series is drawn from an AR(1) process with $\alpha = 0$, i.e., random walk, and the other half is drawn from an AR(1) process with $\alpha > 0$.

The transition probabilities for each part, $p_{i,j}^k$, $k = 1, 2$ and also for the full period $p_{i,j}^f$, $i, j \in \{U, D\}$ are calculated by using the procedure explained above. Note that since $\alpha = 0$, all the transition probabilities for the first part are equal to $1/2$, $p_{i,j}^1 = 1/2$; $i, j \in \{U, D\}$. However, for the second part, $p_{UU}^2 = p_{UU}(\alpha)$. Therefore, the transition probability for observing two upward movements for the full period will be estimated as $p_{UU}^f = (1/2 + p_{UU}(\alpha))/2$. Then the time-homogeneity test statistic given in Eq. (10) yields

$$-2 \ln(A) = 2 \left[\sum_{i,j} n_{i,j}^1 \ln \frac{p_{i,j}^1}{p_{i,j}^f} + \sum_{i,j} n_{i,j}^2 \ln \frac{p_{i,j}^2}{p_{i,j}^f} \right], \quad (19)$$

$$i, j \in \{U, D\}.$$

Let us approximate the number of observations for each transition $n_{i,j}^k$ proportional to these transition probabilities. Note that since $P[y(t) > \bar{y}] = 1/2$, we expect to see half of the observations from state U and the other half from state D . Then, $n_{i,j}^k = (N/4)p_{i,j}^k$. Finally, Eq. (10) can be written as a function of α as

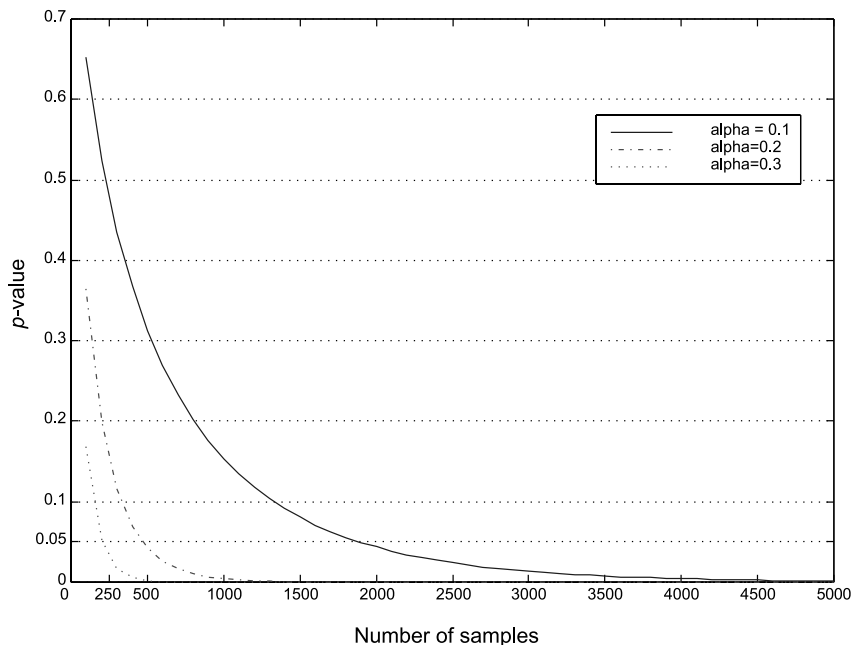


Fig. 4. p -values for the Markov chain random walk test.

$$\begin{aligned}
-2 \ln(A) = & \frac{N}{2} \left[\ln \frac{1}{(p_{UU}(\alpha) + 1/2)} \right. \\
& + p_{UU}(\alpha) \ln \frac{2p_{UU}(\alpha)}{(p_{UU}(\alpha) + 1/2)} + (1 \\
& \left. - p_{UU}(\alpha)) \ln \frac{2(1 - p_{UU}(\alpha))}{3/2 - p_{UU}(\alpha)} \right],
\end{aligned}$$

where $p_{UU}(\alpha)$ is evaluated from Eq. (17). The p -value of the test statistic can be determined from a χ^2 distribution with two degrees of freedom. Fig. 5 shows the p -value for different values of the difference between AR(1) parameters, $|\Delta\alpha| = |0 - \alpha| = \alpha$ and for different sample sizes.

It is clear from Fig. 5 that when the discrete jump in the parameter that generates the structural break is close to zero, e.g., when $\Delta\alpha = 0.1$, one needs close to 5000 observations in order for the Markov chain test to detect the structural break. Even in the case of $\Delta\alpha = 0.2$, the minimum sample size to guarantee that Markov test attains high statistical power is close to 1100. Therefore, for small discrete jumps in the parameter Markov chain test is likely to lead to a Type-II error when the sample size is small.

However, as $\Delta\alpha$ increases, it becomes easier for the Markov chain test to detect this difference. For example, when $\Delta\alpha = 0.3$ and 0.4, the rejection of the null of homogeneity at the 10% level of significance is guaranteed if the sample size is close to 500 and 250 observations, respectively. As the AR(1) parameter jump increases further towards 1, then a sample size of 100 or less will be sufficient for Markov chain test to detect the structural break from random walk to an AR(1) process. This observation implies that when the Markov chain test is employed on a short sequence, the test results might be distorted due to undetected structural breaks in the time series. Finally, it is interesting to note that the minimum sample size for the Markov chain to reject the incorrect null of time-homogeneity is inversely related to the square of $\Delta\alpha$ approximately.

5. Empirical evaluation of the size and power of the Markov chain test via Monte Carlo simulation

Above we invoked the law of large numbers to analytically evaluate the statistical properties of

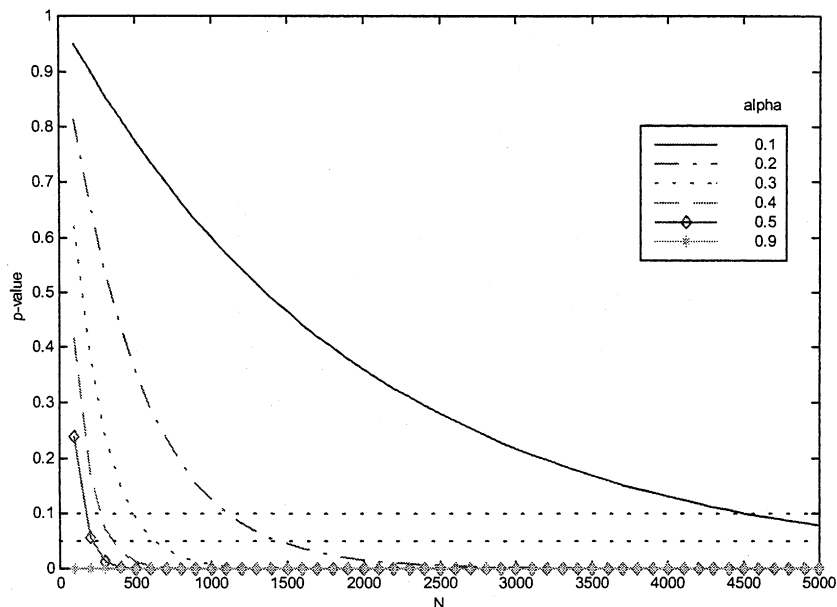


Fig. 5. p -values for the Markov chain time-homogeneity test.

Markov chain time-dependence and homogeneity tests. However, this assumption is rather restrictive and the results obtained do not tell us much about the statistical properties of these tests in small samples. For this reason, in this section we use Monte Carlo simulations to undertake an empirical evaluation of the statistical properties of the time-dependence and homogeneity tests under the martingale null hypothesis as well as the integrated AR(1) and AR(2) processes. In an earlier study, Guthrie and Youssef (1970) utilized Monte Carlo simulations to evaluate the chi-square tests for the order of a Markov chain empirically. All Monte Carlo simulations are conducted in MATLAB version 5.2.

5.1. Time-dependence

In our size and power calculations for the Markov chain test, we follow the methodology described in Fig. 1 and discussed in full detail in Section 3. Monte Carlo experiments are based on

1000 replications of time series of a given length. In its more general form, each time series $y(t)$ is allowed to follow an AR(2) process, which can be written as

$$y(t) = \alpha_1 y(t-1) + \alpha_2 y(t-2) + \varepsilon(t), \quad t = 2, 3, \dots, \quad (20)$$

where α_1 and α_2 are scalars and $\varepsilon(t)$ is the noise term that has a normal distribution with mean 0 and variance σ^2 . We generate $y(0)$ and $y(1)$ as uniform random numbers and use the recursive Eq. (20) to generate the time series. The general AR(2) format allows us to consider AR(1) by setting $\alpha_2 = 0$, as well as random walk processes by setting $\alpha_1 = \alpha_2 = 0$. We consider auto-regressive process in stock returns only and not in stock prices, because Lo and MacKinlay (1988) results showed that AR(1) in stock returns is the most plausible alternative to RWH.

We start with the empirical size calculations of the time-dependence test with the maintained hypothesis of time-homogeneity. Table 2 reports empirical size calculations for a nominal test size

Table 2
Empirical size of the Markov chain time-dependence test (at the 5% level in 1000 Monte Carlo replications)^a

σ^2	N	Two states				Three states			
		Estimated order of dependence				Estimated order of dependence			
		0	1	2	>2	0	1	2	>2
$\alpha_1 = 0, \alpha_2 = 0$									
0.5	50	957	39	2	2	935	59	6	0
	100	957	43	0	0	954	40	5	1
	250	950	46	3	1	957	39	4	0
	500	942	57	1	0	950	47	3	0
	1000	949	48	2	1	948	47	4	1
1	50	943	49	6	2	932	62	6	0
	100	946	51	2	1	944	52	4	0
	250	946	50	4	0	960	37	3	0
	500	944	50	6	0	950	46	3	1
	1000	947	51	2	0	943	53	4	0
1.5	50	955	44	0	1	937	56	7	0
	100	952	46	2	0	953	43	4	0
	250	934	62	4	0	956	44	0	0
	500	953	45	2	0	949	49	2	0
	1000	949	48	3	0	953	44	3	0

^a H_0 : Random walk (not time-dependent) – H_1 : Time-dependent. 1000 replications of time series $y(t) = \alpha_1 y(t-1) + \alpha_2 y(t-2) + \varepsilon(t)$, $t = 2, 3, \dots, N$, where α_1 and α_2 are scalars and $\varepsilon(t)$ is the noise term that has a normal distribution with mean 0 and variance σ^2 . Columns 0, 1, and 2 report the number of replications for which the respective order of dependence cannot be rejected against a higher-order, whereas the >2 column reports the number of replications for which the order of dependence is higher than 2.

of 5% based on 1000 replications of time series with 50, 100, 250, 500 and 1000 observations each. Columns with labels 0, 1, 2, >2 show the number of time series (out of 1000) for which the estimated order of the Markov chain is 0, 1, 2 and >2, respectively. For example, out of 1000 time series generated with 250 observations and $\sigma = 0.5$, the two-state Markov chain test finds the correct order of zero for 950 of them. That is, the empirical size of the test is exactly equal to its nominal size. Table 2 provides more information than just the size of the test. In 46 out of 1000 replications, the Markov chain test rejects the random walk incorrectly and finds an estimated order of one. Only for 3 and 1 time series, respectively, the Markov chain estimates the order to be two or greater.

Irrespective of the sample size and number of states used, empirical size of the Markov chain test is very close to its nominal value, 5%. With the exception of a few cases in small samples, the size distortion, the difference between the empirical and nominal size, is less than 1%. It is clear from Table 2 that the size performance of the Markov chain as a time-dependence test does not depend on the magnitude of the noise in the data. As we increase the variance of the random walk process from 0.5 to 1.0 and 1.5, there is no clear pattern of change in size distortion, one way or the other.

Our size calculations compare well with that of VR and Box-Pierce Q -tests under homoscedastic increments reported in Lo and MacKinlay (1989, Table 1a). In both tests, size distortion varies with the aggregation interval (also called as, return horizon) given the sample size. With homoscedastic stock returns, the empirical size of the VR-test statistic increases slightly above the nominal test size 0.05 as the aggregation increases from two weeks to half of the sample size. However, the empirical size of the VR-test statistic is not distributed evenly on both tails of the test. Most of the rejection is obtained in the upper tail. Being a one-sided test, Q -test does not have this problem. However, its empirical size falls quickly below 5% as the aggregation interval is increased from two weeks to half of the sample size considered. Markov chain test does not suffer from either of these problems and therefore is a more reliable test in

terms of the test size compared to the other two tests.

Next we analyze, the power of Markov chain test under integrated AR(1) and AR(2) alternatives to random walk process. Power calculations are reported in Table 3. In power calculations, we count the number of times the random walk null hypothesis is rejected out of 1000 replications, given that the true process is either an integrated AR(1) or integrated AR(2). The power of the Markov chain time-dependence test is quite low for low values of α_1 . For example, with 1000 observations and $\alpha_1 = 0.1$, two-state Markov chain test rejects the random walk only 515 (1000 minus the number of times estimated order is equal to zero) times out of 1000 replications. Even as the sample size approaches to asymptotic levels, Markov test is not able to reject the random walk. This result confirms the analytical findings of Section 4.

The power of the Markov chain test increases with the AR(1) parameter. Statistical power of the test reaches close to 100% as the sample size reaches to 500 and 1000 observations. For parameter values greater than 0.3, 97.5% rejection rate is reached at sample size 250. However, with parameter values quite plausible for stock returns, such as 0.2 and 0.3, it is still difficult to reject the RWH at smaller sample sizes when it is incorrect. This is not a problem for the Markov chain test only. Lo and MacKinlay's power calculations for various test statistics indicate low power for low parameters values. Apparently, when the AR(1) parameter is close to 0, all tests have difficulty distinguishing between the AR(1) and random walk processes.

Despite these problems, even at 0.2 and 0.3 AR(1) parameter values the power of the Markov chain test increases very rapidly with the sample size. For example, for $\alpha_1 = 0.3$, the power of the Markov chain test with two states reaches to 86.5% level at 250 observations, 98.5% at 500 observations and 100% at 1000 observations. As the AR(1) parameter increases to 0.5, the power of the test reaches to 90% at 100 observations.

We compare the power of Markov chain test with the power of the VR, Box-Pierce Q - and Dickey-Fuller t -tests reported in Lo and

Table 3

Power of the Markov chain time-dependence test (at the 5% level in 1000 Monte Carlo replications of AR(1) or AR(2) series)^a

$\sigma^2 = 1$			Two states Estimated order				Three states Estimated order			
α_1	α_2	N	0	1	2	>2	0	1	2	>2
0.1	0	50	940	53	7	0	934	64	2	0
		100	914	83	3	0	942	50	8	0
		250	834	162	4	0	879	113	7	1
		500	714	274	12	0	757	228	14	1
		1000	485	482	31	0	519	462	17	26
0.2	0	50	877	113	10	0	896	93	11	0
		100	760	223	16	1	819	157	23	1
		250	497	478	24	1	549	410	31	10
		500	184	778	28	0	180	782	37	1
		1000	23	932	44	1	18	918	58	6
0.3	0	50	763	217	19	1	785	188	27	0
		100	577	400	22	1	592	377	37	1
		250	135	826	39	0	149	807	38	6
		500	15	925	60	0	1	944	48	7
		1000	0	933	60	7	0	939	60	1
0.4	0	50	608	364	27	1	634	334	32	0
		100	296	673	30	1	298	643	56	3
		250	24	920	55	1	16	912	61	11
		500	0	907	91	2	0	924	69	7
		1000	0	839	148	13	0	935	58	7
0.5	0	50	405	548	47	0	388	571	41	0
		100	110	837	46	7	98	827	73	2
		250	3	921	73	3	1	927	59	13
		500	0	827	164	9	0	921	72	7
		1000	0	681	299	20	0	868	124	8
0.8	0	50	45	876	73	61	36	935	29	0
		100	0	842	101	17	1	904	95	0
		250	0	580	382	37	0	785	209	6
		500	0	243	648	109	0	625	364	11
		1000	0	26	722	252	0	294	647	59
0.2	0.2	50	813	172	14	1	812	153	35	0
		100	697	251	28	4	713	227	56	4
		250	343	360	286	11	353	426	183	38
		500	84	201	663	52	72	332	525	71
		1000	6	22	910	62	1	30	898	71
0.2	0.6	50	581	197	207	15	559	231	210	0
		100	317	44	571	68	274	49	657	20
		250	32	0	835	133	27	0	871	102
		500	0	0	798	202	0	0	871	129
		1000	0	0	577	423	0	0	849	151

^a H_0 : Random walk (not time-dependent) – H_1 : Time-dependent. 1000 replications of time series $y(t) = \alpha_1 y(t-1) + \alpha_2 y(t-2) + \varepsilon(t)$, $t = 2, 3, \dots, N$, where α_1 and α_2 are scalars and $\varepsilon(t)$ is the noise term that has a normal distribution with mean 0 and variance σ^2 . Columns 0, 1, and 2 report the number of replications for which the respective order of dependence cannot be rejected against a higher-order, whereas the >2 column reports the number of replications for which the order of dependence is higher than 2.

MacKinlay (1989) for the same alternative hypothesis, AR(1) in first differences or equivalently, ARIMA(1,1,0) in levels, with $\alpha_1 = 0.2$. Power of the Markov chain test is higher than the Dickey-Fuller t -test for all sample sizes considered. The comparison with the VR and the Box-Pierce Q -tests results in rather mixed conclusions. Both tests are based on auto-correlations of the return series and have to be calculated for different aggregation intervals. Unfortunately, their power is not constant as the aggregation interval is increased for a given sample size. For example, for a sample size of 256, the power of the VR declines from 89% to 10% as the aggregation interval at which the VRs are calculated is increased from 2 to 64 (see Lo and MacKinlay, 1989, Table 5(e)). The same downward trend in the power of these tests is observed for larger samples as well. For example, for aggregation intervals equal to 32 weekly observations, the power of the VR-test is 0.30 and 0.51 for samples with 512 and 1024 observations, respectively.

The rapid decline of the power with aggregation interval is not a desirable property. After all, serial correlation in stock returns does not have to be relevant for very short return horizon such as two weeks. It is possible to observe departure from random walk at longer return horizons such as 16 or 32 weeks, but apparently neither the VR, nor the Box-Pierce Q -tests have much statistical power to reject the random walk null hypothesis when the underlying process is an integrated AR(1) with $\alpha_1 = 0.2$. As the Markov chain test is not sensitive to aggregation interval, it is preferred to VR and Box-Pierce Q -tests in the case of an integrated AR(1) alternative.

Even though it has higher power than other tests against a RW alternative, a Markov chain test detects the correct order of dependence under very strict circumstances. In general, it is likely to reject the true order of dependence in favor of a lower order of dependence when the sample size is small. For AR(1) and AR(2) parameters close to zero, it is able to find the true order of dependence only for large number of observations. As the AR(1) or AR(2) parameters get closer to 1.0, it becomes biased towards higher-order of dependence.

5.2. Time-homogeneity

In this subsection, we repeat the size and power calculations of the previous subsection for the time-homogeneity test. First, we discuss the empirical size performance of the test, the results of which are presented in Table 4.

In this experiment, we randomly generate 1000 replications of time-homogeneous series and apply the Markov chains to these series to test whether the Markov chain will reject the null of time-homogeneity when it is correct. As described earlier in the paper, we divide our sample into 2, 4 and 8 equal sub-intervals and test whether the transition probability matrix of the Markov chain for these sub-intervals are statistically equal to the one for the full sample. We report the number of rejections and the number of inconclusive results (under the columns denoted by NTH and *, respectively).

Our Monte Carlo experiment results clearly show that Markov chain time-homogeneity test suffers from size distortion problem for all sample sizes and the underlying time series process considered. To start with, when the time series follow random walk throughout the full sample and the null is tested against a two sub-interval alternative, the empirical size of the time-homogeneity test with two states is smaller than its nominal size, 5%. For both small and large sample sizes, the size distortion is close to 3%. The size distortion is even larger in the case of a three-state Markov chain time-homogeneity test and increases with the sample size.

For small values of the AR(1) parameter such as 0.2, the size distortion of a two-state Markov chain time-homogeneity test against one structural break alternative gets closer to the nominal size as the number of observations increases. However, the size distortion problem intensifies as the AR(1) and AR(2) parameters are increased further up or as the number of sub-intervals that are structurally different from each other is increased to 4 and 8. The size distortion is an even more significant problem for the three-state time-homogeneity test. With this result from the Monte Carlo experiments we can conclude that Markov chain test for time-homogeneity does not retain its asymptotic properties in a finite sample setting.

Table 4

Empirical size of the Markov chain time-homogeneity test (at the 5% level in 1000 replications of time-homogeneous series of length N)^a

S	Two states						Three states					
	2		4		8		2		4		8	
	NTH	*	NTH	*	NTH	*	NTH	*	NTH	*	NTH	*
$\alpha_1 = 0, \alpha_2 = 0$												
100	15	0	8	0	5	1	7	1	45	0	167	0
250	19	0	10	0	3	0	9	0	13	0	47	0
500	20	0	9	0	6	1	6	1	10	0	12	0
1000	13	0	2	0	11	1	5	0	3	2	5	0
$\alpha_1 = 0.2, \alpha_2 = 0$												
100	28	1	23	3	34	3	38	1	162	1	322	0
250	33	1	52	1	48	1	68	2	59	5	449	4
500	45	2	65	2	68	0	75	5	86	9	123	7
1000	54	1	78	0	80	2	82	8	121	5	114	2
$\alpha_1 = 0.8, \alpha_2 = 0$												
100	179	14	341	17	477	8	930	0	1000	0	1000	0
250	176	37	362	24	587	39	713	4	978	4	994	6
500	200	112	420	112	687	98	518	17	901	24	988	12
1000	185	250	358	252	624	254	704	68	786	68	931	61
$\alpha_1 = 0.2, \alpha_2 = 0.2$												
100	85	1	115	3	150	5	86	6	318	4	543	6
250	132	18	203	15	329	9	195	56	266	42	577	39
500	178	46	313	40	558	41	534	69	587	64	637	66
1000	213	71	395	74	612	60	754	62	890	62	893	64
$\alpha_1 = 0.2, \alpha_2 = 0.6$												
100	323	61	596	72	710	72	673	22	731	14	905	21
250	316	114	620	108	826	113	881	96	854	116	856	112
500	278	199	557	194	754	204	847	132	862	138	844	156
1000	235	385	613	144	811	142	758	166	824	176	824	176

^a H_0 : Time-homogeneous – H_1 : Not time-homogeneous. S stands for the number of sub-intervals in the alternative hypothesis. NTH column reports the number of rejections of the null hypothesis, whereas * column reports the number of Monte Carlo experiments inconclusive test results.

Unlike the two-state Markov chain, when we use three states the sample size and the number of equal-length sub-periods do matter. Number of times the null of time-homogeneity is rejected increases with an increase in the number of sub-periods, but decreases with an increase in the sample size. These results are expected. As the number of sub-periods increases from 2 to 4, to 8 and so on, the number of observations in each sub-period and, therefore, the number of degrees of freedom decrease. For example, when the number of sub-periods is 16 and the full sample size is 100, in each sub-sample we have 6 observations to estimate at least 9 transition probabilities. That is why the test

rejects the null of time-homogeneity hypothesis in 984 cases out of 1000 runs. The only 15 acceptances are due to zero-order in which case there is only one parameter to estimate, 1/3.

Finally, we consider the empirical power of the time-homogeneity test when the significance level is set at the 5% level. Table 5(a) and (b) present the power calculations under different assumptions about the AR(1) parameter. Table 5(a) considers a time series, which is created with a structural break either at the 1/3, 1/2 and 2/3 of the total number of observations. In our tests, we divide the series generated in the above fashion into two equal sub-intervals and test for time-homogeneity.

Table 5

Power of the Markov chain time-homogeneity test (at the 5% level in 1000 replications of non-homogeneous series of length N^a)

Actual breakpoint N	Two states						Three states					
	$N/3$		$N/2$		$2N/3$		$N/3$		$N/2$		$2N/3$	
	NTH	*	NTH	*	NTH	*	NTH	*	NTH	*	NTH	*
(a) $\alpha = 0.1, \alpha' = 0$												
100	15	0	18	0	10	0	21	0	14	0	14	0
250	15	0	19	1	18	0	7	3	12	3	14	2
500	18	0	26	0	26	0	11	2	14	0	15	0
1000	29	0	37	0	43	2	11	1	16	0	31	0
$\alpha = 0.2, \alpha' = 0$												
100	19	0	24	0	22	0	19	0	22	0	19	0
250	18	0	31	0	42	0	12	0	31	0	31	0
500	36	0	94	0	86	0	25	0	61	4	69	3
1000	63	0	217	1	177	2	51	2	208	1	175	1
$\alpha = 0.3, \alpha' = 0$												
100	25	0	35	0	42	2	14	0	32	0	50	1
250	47	0	101	1	95	0	27	2	77	3	83	5
500	86	1	287	4	209	4	69	0	287	2	198	7
1000	250	9	695	5	446	4	243	1	735	8	484	5
$\alpha = 0.4, \alpha' = 0$												
100	39	0	70	1	65	1	24	1	76	0	73	4
250	82	2	280	2	187	6	62	2	220	7	220	11
500	194	2	636	5	419	5	190	4	678	7	440	10
1000	543	12	949	14	709	9	614	9	977	7	768	7
$\alpha = 0.5, \alpha' = 0$												
100	59	0	139	3	108	4	61	0	159	4	214	4
250	154	3	513	5	324	12	152	7	531	20	378	21
500	408	4	891	10	604	210	464	12	934	18	664	20
1000	834	13	953	46	877	37	887	14	972	27	921	25
(b) $\alpha = 0.2, \alpha' = 0.4$												
100	58	2	72	2	48	1	119	4	130	2	88	2
250	98	3	146	3	82	1	150	14	169	12	108	11
500	148	6	275	4	146	4	168	8	281	8	169	12
1000	198	5	312	2	186	4	265	8	537	5	277	4
$\alpha = 0.2, \alpha' = 0.8$												
100	357	21	551	21	271	15	747	6	739	11	422	19
250	629	82	869	75	655	42	708	62	873	88	667	76
500	692	244	890	105	823	110	784	176	795	205	825	146
1000	712	224	902	41	877	112	613	387	570	430	677	323
$\alpha = 0.4, \alpha' = 0.6$												
100	130	7	110	8	96	1	412	3	381	0	293	2
250	158	3	219	9	138	6	238	16	244	14	156	14
500	227	15	316	17	198	12	261	11	373	12	203	5
1000	290	12	388	19	253	11	416	19	637	12	380	20
$\alpha = 0.4, \alpha' = 0.8$												
100	271	11	396	18	231	14	772	3	712	6	532	7
250	487	39	712	38	443	21	618	28	753	38	468	28
500	632	135	844	98	682	54	735	82	893	89	722	58
1000	706	129	923	55	761	60	767	220	801	199	858	116

Table 5 (continued)

Actual breakpoint N	Two states						Three states					
	$N/3$		$N/2$		$2N/3$		$N/3$		$N/2$		$2N/3$	
	NTH	*	NTH	*	NTH	*	NTH	*	NTH	*	NTH	*
$\alpha = 0.6, \alpha' = 0.8$												
100	215	17	256	12	151	10	848	2	784	0	705	2
250	241	31	376	32	215	22	568	8	541	8	378	11
500	388	89	567	70	391	71	528	42	670	28	439	27
1000	555	92	671	89	498	82	722	88	853	78	662	59

^a H_0 : Time-homogeneous – H_1 : Not time-homogeneous. Actual breakpoint refers to the data point at which the generated time series is broken into two sub-intervals and tested for time-homogeneity with breakpoint assumed to be in the middle. NTH column reports the number of rejections of the null hypothesis out of 1000 replications, whereas * column reports the number of inconclusive test results.

Consistent with our analytical results presented in Fig. 4, the power of the Markov chain test is rather low for small samples and for small discrete jumps in the data at the structural break point. The power of the test increases to only 22% when the AR(1) parameter is equal to 0.2 in the first half and zero in the second half. The power of the test increases as the discrete jump in the underlying AR(1) parameter increases to 0.3 through 0.5. The statistical power of the test increases above 90% when the AR(1) parameter is equal to 0.4 and the break point is located in the middle of the sample. The results in Table 5(b) confirm the conclusion we reached at Table 5(a), that the power of the Markov chain time-homogeneity test increases with the size of the discrete jump in the AR(1) parameter that generated the break.

To conclude this section, our power and empirical size calculations make it clear that time-homogeneity (or structural break) is difficult to detect with Markov chain test. The Markov chain test for time-homogeneity does not have very desirable finite-sample power and size characteristics. It is not only the Markov chain test that suffers from undesirable power and size calculations as a test of structural break. Diebold and Chen (1996) showed that even asymptotic structural break tests do not perform well.

6. Conclusions

To summarize, in this paper we presented a framework for Markov chain test of time-depen-

dence and time-homogeneity. Based on this methodology, we developed an analytical evaluation of the Markov representation of time series. It is shown that there is a one-to-one correspondence between the AR parameter and the transition probability matrix of the corresponding Markov chain. Second, we derived an analytical measure of the statistical power of the Markov chain time-homogeneity test. It turns out that one needs quite a large sample size in order to have statistical power to correctly reject time-homogeneity when the discrete jump in the AR parameter is small.

We later used Monte Carlo experiments to examine the finite-sample properties of the Markov chain time-dependence and -homogeneity tests. The results showed that under the null hypothesis of an i.i.d. random walk, the empirical size of the Markov chain time-dependence test is close to its nominal value irrespective of the sample size and the number of states. However, unlike the empirical size calculations, sample size matters in power calculations. Under the alternative integrated AR(1) and AR(2) processes, the power of the Markov chain test increases rapidly with the number of observations.

Finally, the Monte Carlo results show that Markov chain test does not have much desirable properties as a test of time-homogeneity. The size of the Markov chain test for time-homogeneity is lower than its nominal value, when the underlying series is a time-homogeneous random walk. For time-homogenous integrated AR(1) process with a low AR(1) parameter, size distortion of the Markov chain test is still low. However, as the sum of

auto-regressive parameters increases, the number of false rejections of the time-homogeneity hypothesis increases as well. The power of Markov chain time-homogeneity test also increases with the sample size and with the increase in auto-regressive parameters. However, Markov chain time-homogeneity test does not perform better than other structural break tests in terms of power calculations.

We believe that the contribution of this study is twofold: First, the presented framework helps one to conduct Markov chain tests for time-dependence and time-homogeneity and allows a statistically meaningful interpretation of the test results. Second, the analytical and empirical evaluation shows the pros and cons of the testing procedure, and, therefore, is of interest to decide whether or not to use Markov chain based tests depending on the properties of the time series and the sample size.

References

- Anderson, T.W., Goodman, L.A., 1957. Statistical inference about Markov chains. *Annals of Mathematical Statistics* 28, 12–40.
- Cootner, P.H., 1962. Stock prices: Random vs systematic changes. *Industrial Management Review* 3, 24–45.
- Diebold, F.X., Chen, C., 1996. Testing structural stability with endogenous breakpoint: A size comparison of analytic and bootstrap procedures. *Journal of Econometrics* 70, 221–241.
- Dryden, M.M., 1969. Share price movements: A Markovian approach. *Journal of Finance* 24, 49–60.
- Fama, E.F., 1965. The behavior of stock market prices. *Journal of Business* 38, 34–105.
- Fama, E.F., French, K.R., 1988. Permanent and temporary components of stock prices. *Journal of Political Economy* 96, 246–273.
- Fama, E.F., Blume, M.E., 1966. Filter rules and stock market trading. *Journal of Business* 39, 226–241.
- Fielitz, B.D., 1975. On the stationarity of transition matrices of common stocks. *Journal of Financial and Quantitative Analysis* 10, 327–339.
- Fielitz, B.D., Bhargava, T.N., 1973. The behavior of stock-price relatives: A Markovian analysis. *Operations Research* 21, 1183–1199.
- Gregory, A.W., Sampson, M., 1987. Testing the independence of forecast errors in the forward foreign exchange market using Markov chains. *International Journal of Forecasting* 3, 97–113.
- Guthrie, D., Youssef, M.N., 1970. Empirical evaluation of some chi-square tests for the order of a Markov chain. *Journal of the American Statistical Association* 65, 631–634.
- Lo, A., 1997. Introduction. In: *Market Efficiency: Stock Market Behavior in Theory and Practice*, An Elgar Reference Collection.
- Lo, A.W., MacKinlay, A.C., 1989. The size and power of the variance ratio test in finite samples a Monte Carlo investigation. *Journal of Econometrics* 40, 203–238.
- Lo, A.W., MacKinlay, A.C., 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies* 1, 41–66.
- Kim, M.J., Nelson, C.R., Startz, R., 1991. Mean reversion in stock prices? A reappraisal of the empirical evidence. *Review of Economic Studies* 58, 515–528.
- McQuenn, G., Thorley, S., 1991. Are stock returns predictable? A test using Markov chains. *Journal of Finance* 46 (1), 239–263.
- Neftci, S.N., 1984. Are economic time series asymmetric over the business cycle?. *Journal of Political Economy* 92, 307–328.
- Niederhoffer, V., Osborne, M., 1966. Market making and reversal on the stock exchange. *Journal of the American Statistical Association* 61, 897–916.
- Osborne, M.F.M., 1959. Brownian motion in the stock market. *Operations Research* 7, 145–173.
- Poterba, J.M., Summers, L.H., 1988. Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics* 22, 27–59.
- Richardson, M., 1993. Temporary components of stock prices: A skeptic's view. *Journal of Business and Economic Statistics* 11, 199–207.
- Sherlaw-Johnson, C., Gallivan, S., Burrridge, J., 1995. Estimating a Markov transition matrix from observational data. *Journal of the Operational Research Society* 46 (3), 405–410.