THE JOURNAL OF FINANCE • VOL. XLVI, NO. 1 • MARCH 1991

Are Stock Returns Predictable? A Test Using Markov Chains

GRANT MCQUEEN and STEVEN THORLEY*

ABSTRACT

This paper uses a Markov chain model to test the random walk hypothesis of stock prices. Given a time series of returns, a Markov chain is defined by letting one state represent high returns and the other represent low returns. The random walk hypothesis restricts the transition probabilities of the Markov chain to be equal irrespective of the prior years. Annual real returns are shown to exhibit significant nonrandom walk behavior in the sense that low (high) returns tend to follow runs of high (low) returns in the postwar period.

SINCE FAMA'S (1970) SURVEY of studies supporting the random walk hypothesis, several departures from a random walk have been documented. One category of departures, reported in two recent papers, consists of serial correlation in long-horizon stock returns. Using variance ratio tests, Poterba and Summers (1988) conclude that a substantial part of the variance in monthly returns is due to a transitory, or predictable, component. The transitory component results in prices that are mean reverting and returns that exhibit negative serial correlation when measured over long horizons. Significant mean reverting behavior is also found in a related test by Fama and French (1988) based on negative first order autocorrelation coefficients of many period returns. Kim, Nelson, and Startz (1991) show that the evidence of mean reversion in these two papers is limited to the 1926 to 1946 period, suggesting that the extraordinary swings in the economy surrounding the Depression and World War II have left an apparent episode of mean reversion.[1]

Several explanations for the observed mean reversion have been suggested. Poterba and Summers (1987) indicate that if fads cause prices to temporarily deviate from their fundamental values, then returns would be negatively serially correlated. Fama and French (1988) suggest that the transitory component is caused by time-varying expected returns and therefore is not in

*McQueen is with the Marriott School of Management, Institute of Business Management, Brigham Young University. Thorley is with the Department of Finance, Graduate School of Business Administration, University of Washington. We are grateful to Haim Levy, J. Michael Pinegar, V. Vance Roley, Andrew Siegel, Simon Wheatley, an anonymous referee, the editor (René Stulz), and seminar participants at the University of Washington and Brigham Young University.

[1]Kim, Nelson, and Startz (1991) additionally show that the evidence of prewar mean reversion in previous tests is overstated due to the assumption of normally distributed returns.

conflict with the efficient market hypothesis. The time-varying return model implies that the negative serial correlation may be explained by "macroeconomic driving variables" correlated with time-varying expected returns. Blanchard and Watson (1982) propose a middle ground between these two explanations, rational speculative bubbles, that incorporates the fads of Poterba and Summers (1988) yet retains the rationality of Fama and French (1988). In the rational speculative bubbles model, investors realize that prices exceed fundamental values, but they believe there is a good probability that the bubble will continue to expand and yield a high return. The high return exactly compensates the investors for the probability of a crash; thus, the model shows the rationality of staying in the market despite the overvaluation.

The variance ratio and regression tests assume linearity, yet the alternatives of fads or rational speculative bubbles suggest the possibility of nonlinear patterns in returns. These theories motivate tests that condition on specific sequences of high or low years. For example, this paper estimates the probability of a below-average year given that *both* prior years are above average. The time series based tests are linear in the sense that they relate *all* the observations in a series to the values of prior observations using one set of parameters. The advantage of the Markov chain approach in this context is that it addresses nonlinearity by allowing the parameters (transition probabilities) to vary depending on a given sequence of prior states.[2] This paper introduces a test of the random walk hypothesis based on the statistical theory of finite state Markov processes, or Markov chains, and provides evidence of nonrandom behavior in *postwar* annual returns. Additionally, this paper develops a randomization experiment for small-sample Markov chain tests and develops a simplification in the incorporation of initial state information.

The model, related to Neftci (1984) and Falk (1986), reduces the random walk hypothesis to restrictions on transition probabilities from one state to another, then tests the restrictions using the likelihood function. Given a time series of returns, we define a Markov chain $\{I_t\}$ by letting one state represent high returns and the other represent low returns. The random walk hypothesis implies that the Markov chain representation is without structure or that the transition probabilities are equal, irrespective of the prior returns. Using the process $\{I_t\}$, transition probability estimates for annual returns are calculated and used to form likelihood ratio tests of the equal transition probabilities constraint imposed by the random walk hypothesis. Unlike the tests of Poterba and Summers (1988) and Fama and French (1988), the tests reported in this paper are *nonlinear* in the sense that they

[2]An additional benefit of the Markov chain methodology is that it allows for tests of the random walk under a different set of assumptions than are traditionally needed. For example, the Markov chain tests do not require annual returns to be normally distributed although they do require the Markov chain to be stationary. Markov chain stationarity is defined as constant transition probabilities over time. One cost of modeling returns with Markov chains is the information that is lost when continuous valued returns are divided into discrete states.

focus on periods of established trends and test whether a low return is more likely after observing a sequence of 2 high return years or a sequence of 2 low return years. Furthermore, the results of the Markov chain tests in this paper reject the random walk for *postwar* annual returns based on the relationship to the prior 2 years only, whereas Poterba and Summers (1988) and Fama and French (1988) reject the random walk in the *prewar* period for return horizons of 3 or more years.

Our paper is organized as follows: Section I adapts the Markov chain model to tests of the random walk hypothesis. Section II describes the data, then reports the results of likelihood ratio tests for annual and weekly returns. Section III addresses the statistical issues of using Fisher's Exact Test, relying on asymptotic tests, and ignoring initial state information. The paper's conclusions are presented in Section IV.

I. The Markov Chain Model

In this section, returns are modeled as a two-state Markov chain, and the random walk hypothesis is shown to impose a restriction on the transition probabilities. The separation of returns into two states is similar in nature to the tests for speculative bubbles proposed by Blanchard and Watson (1982) who employ a "Runs" test on gold prices and by Evans (1986) who employs a "Median" test on exchange rates. Markov chain tests were applied to daily stock returns in several early papers and recently have been applied to other economic time series.[3]

In this paper the Markov chain methodology is applied to annual stock returns. Let R_t denote the t^{th} return in a stationary time series of T returns, and let $\{I_t\}$ be given by:

$$I_t = \begin{cases} 1 & \text{if } R_t > \overline{R} \\ 0 & \text{if } R_t < \overline{R} \end{cases}$$

where $\overline{R}$ is a measure of expected returns used to divide the data into high and low return years. In the tests that follow, the rolling average return of

[3]Niederhoffer and Osborne (1966) use Markov chains to show some nonrandom behavior in transaction to transaction ticker prices resulting from investors' tendency to place buy and sell orders at integers (23), halfs ($23\frac{1}{2}$), quarters, and odd eighths in descending preferences. Dryden (1969) applies Markov chains to U.K. stocks which, at the time, were quoted as rising, falling, or remaining unchanged. Fielitz and Bhargava (1973) and Fielitz (1975) show that individual stocks tend to follow a first order, or higher, Markov chain for daily returns; however, the process is not stationary. Samuelson (1988) uses a first order Markov chain to explore the implications of mean regressing equity returns.

A two-state Markov chain is used by Turner, Startz, and Nelson (1989) to model changes in the variance of stock returns, and Cecchetti, Lam, and Mark (1990) show that if economic driving variables follow a Markov chain process, then the negative serial correlation found in long horizons can be consistent with an equilibrium model of asset pricing. Markov chains have been used to model other asset markets, for example, Engle and Hamilton (1990), Gregory and Sampson (1987), and Hamilton (1989).

the prior 20 years is used as the measure of ex ante expectations. The sensitivity analysis discussed in Section II indicates that the test results are robust with respect to the measure of $\overline{R}$.

The derived series $\{I_t\}$ is a two-state Markov chain that represents high returns as ones and low returns as zeros. Employing a second order Markov chain, the transition counts and transition probabilities form the following 4 by 2 matrices:[4]

Transition Count Matrix

Previous States		Current State 0	Current State 1
0	0	N_{00}	M_{00}
0	1	N_{01}	M_{01}
1	0	N_{10}	M_{10}
1	1	N_{11}	M_{11}

Transition Probability Matrix

Previous States		Current State 0	Current State 1
0	0	λ_{00}	$1 - \lambda_{00}$
0	1	λ_{01}	$1 - \lambda_{01}$
1	0	λ_{10}	$1 - \lambda_{10}$
1	1	λ_{11}	$1 - \lambda_{11}$

where $N_{00}, \ldots, M_{11}$ are the number of observed occurrences of the associated transitions. For example, N_{01} is the number of observations of state sequence 0 1 0, and M_{01} is the number of observations of state sequence 0 1 1, where the transition probabilities are defined as:

$$\lambda_{00} = P[I_t = 0 \mid I_{t-2} = 0, I_{t-1} = 0]$$
$$\lambda_{01} = P[I_t = 0 \mid I_{t-2} = 0, I_{t-1} = 1]$$
$$\lambda_{10} = P[I_t = 0 \mid I_{t-2} = 1, I_{t-1} = 0]$$
$$\lambda_{11} = P[I_t = 0 \mid I_{t-2} = 1, I_{t-1} = 1]$$

The maximum likelihood estimates (MLE) of the four unknown parameters $\Lambda = [\lambda_{00}\ \lambda_{01}\ \lambda_{10}\ \lambda_{11}]'$ are found by setting the partial derivatives of the log likelihood function equal to zero and solving for the four parameters in terms

[4]The tests that follow do not assume that returns follow a two-state second-order Markov chain but rather utilize the fact that *if* returns follow a random walk, then the transition probabilities must be without structure. The choice of a second-order representation for the annual returns is appropriate for two reasons: 1) it coincides with the "long horizon" nature of the Fama and French (1988) and Poterba and Summers (1988) tests, and 2) it dominates either a first- or third-order chain in Bayesian Information Criterion and Likelihood Ratio tests. The results of these chain order tests are available from the authors.

Kroll, Levy, and Rapoport (1988), using experimental data, find that investors behave as if stock returns have a two-period memory. Specifically, statistically knowledgeable students, receiving monetary payoffs contingent on performance, show tendencies to switch from risky asset A to risky asset B after A out-performed B for two consecutive trials. This pattern of behavior is evident even though the students know the two returns are independent and normally distributed.

Neftci (1984) and subsequently Falk (1986) use a second-order, two-state Markov chain model to test for asymmetries in the business cycle. The null hypothesis and test statistics used in this paper differ from Neftci; consequently, the notation differs also.

of transition counts. The log likelihood function is:

$$L(S_{\mathrm{T}}, \Lambda', \pi) = \log \pi + \sum_{ij=0\ 0}^{1\ 1} N_{ij}\log \lambda_{ij} + M_{ij}\log(1 - \lambda_{ij}) \tag{1}$$

where S_{T} is the realization of $\{I_t\}$ and π is the probability of the initial two states. As noted by Neftci (1984), π can be ignored if the sample size T is large, so that the information contained in π becomes negligible. Incorporating the initial states requires an iterative solution to Λ, whereas ignoring π allows one to solve the first-order conditions and obtain the trivial result that the maximum likelihood estimates equal the sample probabilities. For example, the maximum likelihood estimate of λ_{11} is $N_{11}/(N_{11} + M_{11})$.[5] Section III.*B* shows that the postwar sample of returns is sufficiently long that excluding π does not change the outcome of the tests; therefore, in the results that follow the initial states are ignored.

The random walk hypothesis requires the probability of state 0 or state 1 to be the same for any prior two-state sequence. In terms of the model, the null hypothesis of a random walk is that the transition probabilities are all equal, $\lambda_{00} = \lambda_{01} = \lambda_{10} = \lambda_{11}$. In particular, the fad and rational bubbles alternatives suggest that the probability of state 0 should be greater for the prior state sequence of 1 1 than 0 0, $\lambda_{00} < \lambda_{11}$. Because the interesting alternatives imply that price reversions or bubble bursts occur after established trends or runs, the tests that follow focus primarily on the null hypothesis, $\lambda_{00} = \lambda_{11}$.

The likelihood ratio test (LRT) is used to test both null hypotheses implied by a random walk; Hypothesis 1, $\lambda_{00} = \lambda_{11}$, and Hypothesis 2, $\lambda_{00} = \lambda_{01} = \lambda_{10} = \lambda_{11}$:

$$\mathrm{LRT} = 2[L(\hat{\Lambda}) - L(\tilde{\Lambda})] \sim \chi_n^2 \tag{2}$$

where $L(\hat{\Lambda})$ is the log likelihood function evaluated at the *unconstrained* MLE of the parameters and $L(\tilde{\Lambda})$ is the log likelihood function using the MLE of the *constrained* parameters.[6] The likelihood ratio test is asymptotically distributed χ_n^2 with n degrees of freedom equal to the number of restrictions. However, results of a randomization experiment reported in Section III.*A* show that the LRT rejects the null too often with only 41 years of postwar data; consequently, the marginal significance we report for the LRT is based on the simulations.

In addition to the LRT's, confidence regions are plotted for λ_{00} and λ_{11} to illustrate the evidence of nonlinear serial dependence. The boundary of the

[5]Ignoring the initial states, the MLE of λ_{ij}, $N_{ij}/(N_{ij} + M_{ij})$ and its asymptotic variance, $\sigma^2(\lambda_{ij}) = \lambda_{ij}(1 - \lambda_{ij})/(N_{ij} + M_{ij})$, are related to the mean and variance of the binomial distribution. The simulation results in Section III.*A* imply that the asymptotic variances reported in Table II and the confidence regions in Figures 1 and 2 are slightly understated due to the small sample size.

[6]In addition to likelihood ratio tests, Wald tests and Lagrange multiplier tests are performed as explained in Section III.*A*.

confidence regions are the locus of λ_{00} and λ_{11} pairs that solve the following equation based on the likelihood ratio test of the hypothesis that $\lambda_{00} = \tilde{\lambda}_{00}$ and $\lambda_{11} = \tilde{\lambda}_{11}$:

$$2[L(\hat{\Lambda}) - L(\Lambda)] = \chi_2^2(\alpha) \tag{3}$$

where α is the selected confidence level.[7] All points within the confidence region may represent the true λ_{00} and λ_{11} with probability α. The random walk hypothesis requires that the probability of state 0 or state 1 should be the same for any prior two-state sequence. If the confidence region plots completely above (below) the 45-degree line (with λ_{00} measured on the horizontal axis), negative (positive) serial dependence is indicated and the random walk is rejected. The confidence regions found using equation (3) are less powerful tests of the random walk than the likelihood ratio tests of (2). The LRT using equation (2) is a direct test of Hypothesis 2 with only one constraint (distributed χ^2 with *one* degree of freedom), whereas the confidence regions found using (3) employ two constraints by solving for all combinations of λ_{00} and λ_{11} that satisfy the equation (distributed χ^2 with *two* degrees of freedom). Nevertheless, the location of the confidence regions with respect to the 45-degree line provides an illustration of the serial dependence of returns.

II. Test Results

Tests for negative serial dependence in postwar annual returns are performed on both real and excess continually compounded returns for equally- and value-weighted portfolios of all New York Stock Exchange (NYSE) stocks for the calendar years 1947 to 1987. Annual continuously compounded nominal returns are created from 1-month returns from the Center for Research in Security Prices (CRSP). To create real returns, continuously compounded annual inflation rates generated from Ibbotson and Sinquefield's (1988) monthly inflation series (based on the Consumer Price Index without seasonal adjustment) are subtracted from the nominal rates. To create excess returns, an annual continuously compounded risk-free rate of return is created from the CRSP RISKFREE file of 1-month Treasury Bill rates and subtracted from the nominal rates.

In addition to the annual tests, we confirm the Lo and MacKinlay (1988) finding of positive serial correlation in weekly returns using the Markov chain model.[8] The weekly return data are created from the CRSP daily return index for equally- and value-weighted portfolios of NYSE and Ameri-

[7]Numerical methods are required to find each λ_{00} and λ_{11} pair that solves equation (3). The untested parameters λ_{10} and λ_{01} are assumed to be equal to their maximum likelihood estimates.

[8]Lo and MacKinlay (1988) using variance ratio tests on weekly returns find evidence of positive serial correlation in postwar data. After developing a simple model of nontrading, they conclude that the evidence of positive serial correlation cannot be attributed solely to infrequent trading.

can Stock Exchange common stocks (Monday open to Friday close). The sample runs from Monday, July 2, 1962 to Monday, December 21, 1987. Subsample tests are also performed by splitting the data into two equally sized subsamples.

Table I shows the summary statistics for the annual portfolio returns. The negative first and second-order autocorrelation coefficients suggest negative serial correlation in annual returns. If returns are serially uncorrelated, Q will be asymptotically χ_n^2; however, if a number of the sample autocorrelations are not close to zero, Q will be inflated. The Ljung-Box (1978) portmanteau tests at lag 6 indicate serial correlation in both the equally- and value-weighted portfolios at the 5 percent significance level.[9] The significant negative second-order autocorrelations and the Q statistics are both linear-based tests of serial correlation that portend the rejection of the random walk found using the nonlinear Markov chain tests.

Table II reports the Markov chain transition counts and MLE of the transition probabilities that result from dividing the annual return series into high and low returns. High (low) returns result when the annual real return is greater (less) than the average return of the previous 20 years. The unconstrained point estimates for λ_{00} and λ_{11} in Table II clearly indicate the presence of negative serial dependence in annual returns. For example, the real returns for the equally-weighted portfolio had a low year only 2 out of 8 times when the preceding 2 years were also low. However, the portfolio had 9 out of 10 low years when the preceding 2 years were high. In other words, over the 41 years since World War II the probability of seeing a low year after a sequence of 2 high years, 0.90, is over three times as great as the probability, 0.25, of seeing a low year when the previous 2 years were also low. Reestimating the parameters with the constraint under Hypothesis 1, $\lambda_{00} = \lambda_{11}$, and substituting both the constrained and unconstrained log likelihoods into equation (2) results in a LRT of 8.6. Based on the results of the randomization study in Section III. A, the probability of observing a LRT of 8.6 or larger under the null hypothesis is 1 percent. Consequently, the probability that a return is independent of the previous 2 years' returns can be rejected with 99 percent confidence for the postwar real annual returns on an equally-weighted portfolio. When all four of the prior 2-year patterns of returns are constrained to be equal under Hypothesis 2, $\lambda_{00} = \lambda_{01} = \lambda_{10} = \lambda_{11}$, the random walk is still rejected at the 1 percent marginal significance level. However, in tests using other estimates of expectations $\overline{R}$, the high significance levels of Hypothesis 2's rejection is not as robust as Hypothesis 1's. The robustness of significance levels for Hypothesis 1 over Hypothesis 2 indicates

[9]The choice of n is somewhat arbitrary. Setting n low means that one may not capture what might be significant autocorrelations at high lags. On the other hand, setting n too high results in a dilution of the influence of the first m significant autocorrelations by the remaining $n - m$ autocorrelations. The choice of six autocorrelations is an attempt to balance these two forces. Additionally, since there are only 41 annual observations and since the Q test is asymptotic, the Ljung-Box modification is used because it performs better in small samples.

Table I
Summary Statistics of Annual Post-War Returns for Equally and Value-Weighted Portfolios (January 1947–January 1987)

Mean(x), Med(x), and SD(x) are the sample mean, median, and standard deviation of the variable. ρ_t is the sample autocorrelation at lag t and $s(\rho)$ is the asympotic standard error of the autocorrelations under the null hypothesis of a random walk. $Q(6)$ is the Ljung-Box (1978) portmanteau test statistic for 6 autocorrelations, distributed chi-squared with 6 degrees of freedom, and M.S. is the marginal significance level of the Ljung-Box test. All returns are continuously compounded. Annual real returns are nominal returns created from the CRSP monthly returns less annual inflation created from Ibbotson and Associates' monthly inflation series. Annual excess returns are nominal returns created from the CRSP monthly returns less an annual risk-free rate created from the CRSP RISKFREE series of one month Treasury bills.

Variable (x)	Mean(x)	Med(x)	SD(x)	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	$s(\rho)$	$Q(6)$	M.S.
Real Equally-Weighted	.08413	.12349	.20828	−.009	−.379	.029	.310	−.027	−.264	.156	14.65	.023
Excess Equally-Weighted	.07967	.11720	.20398	−.031	−.377	.043	.320	−.054	−.310	.156	16.44	.012
Real Value-Weighted	.06747	.11251	.16652	−.022	−.313	.126	.358	.091	−.169	.156	13.12	.041
Excess Value-Weighted	.06301	.09613	.16217	−.054	−.315	.144	.384	.085	−.208	.156	15.16	.019

Table II

Maximum Likelihood Estimates and Likelihood Ratio Tests for Annual Returns (1947–1987) of Four Portfolios

The log likelihood function is: $\sum_{ij=0\,0}^{1\,1} N_{ij}\log\lambda_{ij} + M_{ij}\log(1-\lambda_{ij})$. $\hat{\lambda}_{ij}$ = the estimated transition probability of going from a state sequence of $I_{t-2} = i$ and $I_{t-1} = j$ to $I_t = 0$, where i and j equal 1 for high-return years and 0 for low-return years, and $\hat{\sigma}(\lambda_{ij})$ = the associated standard deviation. $N_{00} \ldots M_{11}$ = the number of observed occurrences of the associated transitions. The initial state = the values of I_1 and I_2. WT = the Wald test, LRT = the likelihood ratio test, LMT = the Lagrangian multiplier test, and M.S. = the marginal significance level, which is the probability of obtaining that value of the LRT or higher under the null hypothesis. (NA) designates that the asymptotic variance estimate is degenerate.

Transition Counts		Equally-Weighted Real 0	Equally-Weighted Real 1	Equally-Weighted Excess 0	Equally-Weighted Excess 1	Value-Weighted Real 0	Value-Weighted Real 1	Value-Weighted Excess 0	Value-Weighted Excess 1
	0 0	2	6	2	8	1	5	2	7
	0 1	2	9	3	8	2	9	4	7
	1 0	5	5	7	3	4	6	6	4
	1 1	9	1	8	0	9	3	7	2
Initial State		0 0		0 0		0 0		0 0	

	Equally-Weighted Real	Equally-Weighted Excess	Value-Weighted Real	Value-Weighted Excess
MLE Estimates				
$\hat{\lambda}_{00}$	.250	.200	.167	.222
$\hat{\sigma}(\lambda_{00})$	(.153)	(.126)	(.152)	(.139)
$\hat{\lambda}_{11}$	.900	1.000	.750	.778
$\hat{\sigma}(\lambda_{11})$	(.095)	(NA)	(.125)	(.139)
Hypothesis Tests				
H_1: $\lambda_{00} = \lambda_{11}$				
WT	13.0	(NA)	8.8	8.0
LRT	8.6	14.7	5.8	5.9
LMT	8.1	12.4	5.1	5.6
M.S.[a]	(.01)	(.01)	(.03)	(.03)
H_2: $\lambda_{00} = \lambda_{01} = \lambda_{01} = \lambda_{11}$				
WT	27.5	(NA)	13.6	9.3
LRT	14.0	18.9	10.0	7.1
LMT	12.2	15.7	9.2	6.7
M.S.[a]	(.01)	(.01)	(.03)	(.08)

[a]The M.S. values are from a simulation reported in Section III. *A*.

that prior runs of high or low returns are particularly relevant in the rejection of the random walk hypothesis.

Figure 1 plots the locus of λ_{00} and λ_{11} pairs that solve equation (3) for the 90 and 99 percent confidence levels using the results of the real equally-weighted returns. The 90 percent confidence region plots above the 45-degree line, illustrating the rejection of the random walk and the negative serial relationship in annual returns. Because the confidence regions are constructed using χ^2 values with *two* degrees of freedom, the 99 percent

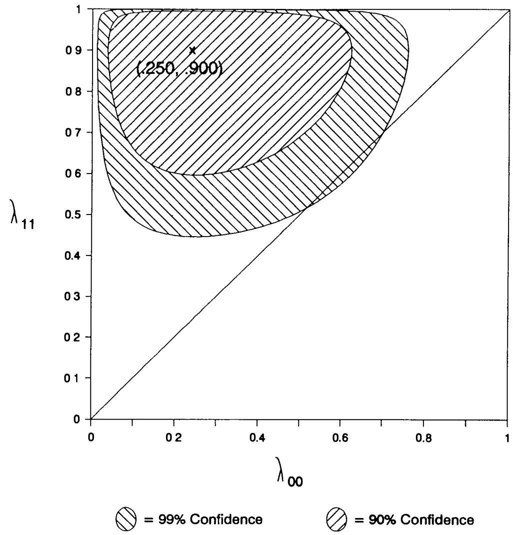

Figure 1. Equally-weighted real annual return (1947-1987) 90 percent and 99 percent confidence regions. λ_{00} = the transition probability that 2 low-return years are followed by a low-return year. λ_{11} = the transition probability that 2 high-return years are followed by a low-return year. X = the position of the unconstrained maximum likelihood estimate of (λ_{00}, λ_{11}). The 45-degree line is the locus of points such that $\lambda_{00} = \lambda_{11}$.

confidence region crosses the 45-degree line even though the direct test of Hypothesis 1 (*one* constraint) is significant at the 1 percent level.

The results for the excess returns of the equally-weighted portfolio and both the real and excess returns of the value-weighted portfolio are also reported in Table II. The equally-weighted excess returns exhibit the same strong rejection of the random walk as the real returns, rejecting both Hypothesis 1 and 2 at the 1 percent significance level. Consistent with the findings of Fama and French (1988), the rejection of the random walk for the value-weighted returns is not as significant as for the equally-weighted returns. However, the value-weighted real portfolio rejects the random walk at the 0.03 significance level for the real returns (both hypotheses) and at the 0.03 and 0.08 significance levels for the excess returns (Hypothesis 1 and 2, respectively). The significant rejection of Hypothesis 1 by the real value-weighted portfolio is illustrated in Figure 2 with the 90 percent confidence ellipse completely above the 45-degree line. The less significant rejection of the random walk by the value-weighted compared to the equally-weighted portfolio is evidenced by the relatively large portion of the 99 percent confidence region that falls below the 45-degree line.

The results of four additional tests refine the strong results of the LRT just discussed. First, the marginal significance levels of Lagrangian multiplier tests and Wald tests confirm the LRT. Second, when the initial states π are incorporated, the evidence against the random walk remains unchanged.[10]

[10]See Section III.*B* for an explanation of incorporating the initial states.

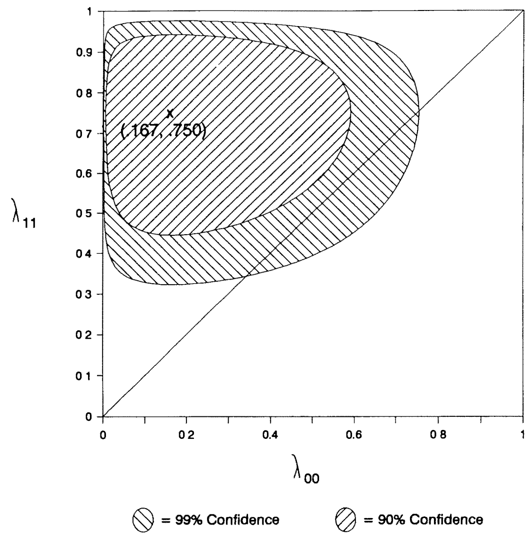

Figure 2. Value-weighted real annual return (1947–1987) 90 percent and 99 percent confidence regions. λ_{00} = the transition probability that 2 low-return years are followed by a low-return year. λ_{11} = the transition probability that 2 high-return years are followed by a low-return year. X = the position of the unconstrained maximum likelihood estimate of (λ_{00}, λ_{11}). The 45-degree line is the locus of points such that $\lambda_{00} = \lambda_{11}$.

Third, in 11 of the 13 times (counting all four portfolios) when a sequence of 0 0 0 or 1 1 1 occurred, the process reversed itself in the next year. In other words, the point estimates from a third-order chain also indicate mean reversion, but observations from the four portfolios are not independent and the limited number of observations make statistical tests of little value.

Fourth, tests based on other measures of $\overline{R}$ consistently reject the random walk hypothesis, indicating that the results are robust with respect to the measure of expectations used to divide returns into high and low states. For example, defining equally-weighted real states for the entire postwar period with a fixed ex ante mean, based on the 20 years from 1927 to 1946, gives identical results to the rolling 20-year mean definition. Sensitivity analysis on the equally-weighted portfolios shows that any fixed measure of $\overline{R}$ between 0 and 15 percent results in the rejection of the random walk (Hypothesis 1) at the 95 percent confidence level or higher. For the value-weighted portfolio, measures of $\overline{R}$ between 0 and 15 percent result in rejections of the random walk at the 90 percent confidence level or higher. This robustness is due to the relatively high variance of annual returns so that most years are high or low by any reasonable measure of expectations.[11] For example, inspection of the equally-weighted real return series reveals that only 2 out

[11]The equally-weighted portfolio is more robust to the measure of expectations used to define high and low states than the value-weighted portfolio. This is partly due to its higher volatility. Results of the sensitivity analysis plus tests based on in-sample median and mean measures of $\overline{R}$ are available from the authors.

of the 41 observations are in the 5 to 10 percent range. Tests using a measure $\bar{R}$ above (below) 0.15 (0.00) have point estimates indicating nonrandom behavior. However, because of the limited number of high (low) states under these more extreme measures of $\bar{R}$, the tests do not have sufficient power to reject the random walk at conventional levels of significance.

Table III Panel A reports the likelihood ratio test statistics for tests of Hypothesis 1 when annual return portfolios are measured for periods other than January to December. Typically, if the return year starts in early-year or late-year months, the evidence against the random walk is significant. On the other hand, if the return year starts in the mid-year months, the evidence against the random walk is not significant. For example, the real equally-weighted returns reject the random walk hypothesis for the 6 returns starting in September through February. The sensitivity of the test results to the starting month is due to the predominance of the January returns and their correlation to prior year and same year returns.[12] Table III Panel B reports the mean January return along with the mean monthly return for February to December months. As reported in several recent studies, January returns are on average higher and more volatile than non-January returns. More importantly, January returns tend to be positively correlated with remaining year returns and negatively correlated with the prior 11 month returns. Noncalendar year periods combine January returns with negatively correlated returns from the prior year and separate January returns from the positively correlated return months that follow. This partitioning tends to negate the negative serial dependence of annual calendar returns found in the Markov chain tests. In general, the results for noncalendar years are stronger for the equally-weighted portfolios than for value-weighted portfolios. This importance of small firms in the rejection of postwar random walks parallels the essentially prewar findings of Fama and French (1988), Poterba and Summers (1988), and Kim, Nelson, and Startz (1991).

Before examining the small sample properties of the LRT, several observations about the rejection of the random walk along with comparisons to existing tests are appropriate. First, the rejection of the random walk is not a rejection of the efficient market hypothesis if expected returns vary through time. However, this explanation only forces the issue one step deeper, requiring an explanation of why expected returns show negative serial correlation.

Second, previous tests have found mean reverting prices in the Depression and World War II period. However, a second-order Markov chain model is not appropriate for the 3- to 5-year pattern of mean reversion found by Poterba and Summers (1988) and Fama and French (1988) in their early subperiods. This early period is characterized by relatively long runs of "good" and "bad" years. For example, the Depression resulted in an unprecedented drop

[12]See Rozeff and Kinney (1976) and Keim (1983) for evidence of the January seasonal. Reinganum (1983) reports evidence that January returns are negatively correlated with the prior months' returns, and Chang and Pinegar (1991) report evidence that January returns are positively correlated with the following months' returns.

Table III

Panel A: Likelihood Ratio Test Statistics of Hypothesis One for Annual Returns Measured from Different Starting Months (1947-1987)[a, b]

Hypothesis 1 states that the random walk restricts the transition probabilities, λ_{00} and λ_{11}, to be equal. λ_{ij} = the transition probability of going from a state sequence of $I_{t-2} = i$ and $I_{t-1} = j$ to $I_t = 0$, where i and j equal 1 for high-return years and 0 for low-return years.

	Equally-Weighted		Value-Weighted	
	Real	Excess	Real	Excess
Jan.-Dec.	8.6**	14.7**	5.8*	5.9*
Feb.-Jan.	3.6*	3.2*	0.2	0.0
Mar.-Feb.	2.2	1.3	0.2	0.2
Apr.-Mar.	2.3	0.9	0.0	0.0
May-Apr.	0.9	0.1	0.0	0.0
June-May	1.8	0.6	1.7	0.0
July-June	0.1	0.0	1.7	0.0
Aug.-July	2.2	0.6	3.7*	3.9*
Sept.-Aug.	4.7*	3.2*	4.6*	3.4*
Oct.-Sept.	4.3*	6.2*	6.5*	5.3*
Nov.-Oct.	3.6*	3.6*	5.8*	8.6**
Dec.-Nov.	9.0**	13.1**	7.2*	7.2*

Panel B: Correlation of January Returns to February-December Period Returns (1947-1987)[c]

$\bar{R}$ is the mean monthly return; S is the sample standard deviation. r(Jan., Post 11) is the correlation coefficient between the return in January and the return over the following 11 months. r(Jan., Pre 11) is the correlation coefficient between the return in January and the return over the preceding 11 months.

	Equally-Weighted		Value-Weighted	
	Real	Excess	Real	Excess
January				
$\bar{R}$	3.95%	3.85%	1.35%	1.25%
S	6.38%	6.43%	5.03%	5.09%
Other 11 months				
$\bar{R}$	0.48%	0.42%	0.54%	0.48%
S	1.68%	1.64%	1.38%	1.30%
r(Jan., Post 11)	0.257	0.285*	0.301*	0.320*
r(Jan., Pre 11)	−0.416**	−0.395*	−0.173	−0.169

[a]The 11 noncalendar year returns use 1988 return data in order to match the 41 year observation length of the January to December return series.

[b]In Panel A: ** and * indicate the likelihood ratio test statistic is significant at the 1 and 10 percent levels, respectively.

[c]In Panel B: ** and * indicate the correlation coefficient is significant at the 1 and 10 percent levels, respectively (see Anderson (1984) for a discussion of sample correlation coefficient significance).

in stock prices over 3 years. Subsequently, the economy recovered, and over the next 4 years surpassed its pre-Depression level. Next, the severe recession in late 1937 and early 1938, followed by the beginning of World War II, resulted in a run of 4 or 5 "bad" years. Finally, the stock market surged again for 4 years as the economy survived Pearl Harbor and engaged in war-time production. These relatively long runs require tests using a third- or fourth-order chain; however, not enough data is available to estimate the 8 or 16 parameters these higher-order chains require.[13]

Third, Fama and French, using the 1941 to 1985 sample, cannot reject the random walk. Similarly, Kim, Nelson, and Startz (1991), using the 1947 to 1986 sample, cannot reject the random walk and, for horizons of 3 or more years, find that the point estimates of postwar multiperiod autocorrelations indicate *positive* serial correlation in returns or "mean aversion" in prices. This disparity is the result of the linear nature of the regression tests used by Fama and French (1988) and Kim, Nelson, and Startz (1991) in contrast to the nonlinear or asymmetric nature of the Markov chain rejection.[14]

Fourth, the conditional sample means reported in Table IV suggest that the serial pattern of returns is economically as well as statistically significant. For example, the real equally-weighted portfolio has a sample mean return of 17.5 percent conditional on 2 prior low return years in contrast to a mean return of -1.7 percent conditional on 2 prior high return years. Hypothesis 3, equal conditional means for prior states of 0 0 and 1 1, can be rejected at the 10 percent level for the equally-weighted real portfolio and at the 5 percent level for the other three portfolios. Hypothesis 4, equal conditional means for all four states, is rejected in only three of the four portfolios and results in higher marginal significance levels than Hypothesis 3, indicating again the importance of prior runs in the rejection of the random walk. The size and significance of the difference in conditional means suggests that the economic importance of predictable returns is not trivial.

In addition to the annual tests, the two-state Markov chain test can be used to reconfirm the positive serial correlation in weekly nominal returns found by Lo and MacKinlay (1988) using variance ratio tests. Table V reports the

[13]Due to the relatively long runs in the prewar period, second-order chain tests for the full (1926–1987) and prewar (1926–1946) periods do not significantly reject the random walk hypothesis, although point estimates suggest negative serial correlation. These results are available from the authors.

[14]The *positive* point estimates of postwar serial correlation for horizons of 3 years or more found by Kim, Nelson, and Startz (1991) are not inconsistent with the *negative* serial relationship found in the second order chain in this paper. Poterba and Summers' (1987) approximation shows that the measure of mean reversion used by Kim, Nelson, and Startz (1991) weighs monthly sample autocorrelation coefficients from 1 to $2k/3$ negatively and higher autocorrelations up to $2k - 1$ positively. For example, using the Poterba and Summers' (1987) approximation, a 3-year regression test incorporates 71 monthly autocorrelations with the first 24 receiving negative weights. Thus, monthly autocorrelations that tend to be negative over the first 24 lags are consistent with both the second order Markov chain finding of an annual negative serial relationship and the positive serial correlation point estimates found in the postwar longer horizon regression tests.

Table IV

Conditional and Unconditional Means and Standard Deviations for the Annual Returns (1947-1987) of Four Portfolios

$\bar{R}_{ij}$ = mean return conditional on a prior state sequence of $I_{t-2} = i$ and $I_{t-1} = j$, where i and j equal 1 for high-return years and 0 for low-return years, and S_{ij} = the associated conditional sample standard deviation. $N_{ij} + M_{ij}$ = the number of observed occurrences of the associated prior two-state sequences. The F-statistic is for a test of equal means based on an analysis of variance (ANOVA), M.S. = the probability of obtaining that value of the F-statistic or higher under the null hypothesis, and df = degrees of freedom.

	Equally-Weighted		Value-Weighted	
	Real	Excess	Real	Excess
Conditional				
$\bar{R}_{00}$	17.5%	18.0%	15.9%	13.7%
S_{00}	29.7%	25.5%	14.0%	15.4%
$N_{00} + M_{00}$	8	10	6	9
$\bar{R}_{01}$	15.4%	15.2%	13.0%	12.1%
S_{01}	12.6%	13.5%	9.2%	9.9%
$N_{01} + M_{01}$	11	11	11	11
$\bar{R}_{10}$	6.6%	2.7%	6.8%	2.4%
S_{10}	24.0%	21.9%	25.0%	22.0%
$N_{10} + M_{10}$	10	10	10	10
$\bar{R}_{11}$	-1.7%	-5.4%	-1.9%	-2.9%
S_{11}	13.5%	12.9%	12.1%	13.5%
$N_{11} + M_{11}$	10	8	12	9
Unconditional				
$\bar{R}$	9.2%	8.5%	7.2%	6.5%
S	21.1%	20.8%	16.9%	16.6%
$N + M$	39	39	39	39
Hypothesis Test				
H_3: $\bar{R}_{00} = \bar{R}_{11}$				
F-statistic	3.36*	5.55*	7.88**	5.91*
M.S. (df = 1,16)	(.09)	(.03)	(.01)	(.03)
H_4: $\bar{R}_{00} = \bar{R}_{01} = \bar{R}_{10} = \bar{R}_{11}$				
F-statistic	1.78	2.91*	2.35*	2.36*
M.S. (df = 3, 35)	(.17)	(.05)	(.09)	(.09)

*Indicates the F-test is significant at the 10 percent level.
**Indicates the F-test is significant at the 1 percent level.

Markov chain model MLE's and LRT statistics for weekly nominal returns on equally-weighted portfolios for the full sample, July 1962 to December 1987, and two subsamples splitting the data at Monday, May 5, 1975. The weekly returns are divided into high and low states based on a fixed ex ante mean weekly return from January 1947 to June 1962. The large number of weekly observations result in similar Wald, likelihood ratio, and Lagrange multiplier tests; thus, only the likelihood ratio is reported. The point estimates of

Table V

Maximum Likelihood Estimates and Likelihood Ratio Tests for Weekly Nominal Returns on a Equally-Weighted Portfolio Based on Fixed Ex Ante Means

The log likelihood function is: $\sum_{ij=0\,0}^{1\,1} N_{ij}\log \lambda_{ij} + M_{ij}\log(1 - \lambda_{ij})$. $\hat{\lambda}_{ij}$ = the estimated transition probability of going from a state sequence of $I_{t-2} = i$ and $I_{t-1} = j$ to $I_t = 0$, where i and j equal 1 for high-return weeks and 0 for low-return weeks, and $\hat{\sigma}(\lambda_{ij})$ = the associated standard deviation. $N_{00} \ldots M_{11}$ = the number of observed occurrences of the associated transitions. The initial state = the values of I_1 and I_2. $\overline{R}$ = the mean weekly return from January 1947 to June 1962. LRT = the likelihood ratio test which is asymptotically distributed χ^2 with 1 and 3 degrees of freedom for H_1 and H_2, respectively, and M.S. = the marginal significance level, which is the probability of obtaining that value of the LRT or higher under the null hypothesis.

Transition Counts		7/62 to 12/87		7/62 to 3/75		4/75 to 12/87	
		0	1	0	1	0	1
	0 0	190	140	122	74	68	65
	0 1	86	157	39	68	47	89
	1 0	140	103	74	33	65	70
	1 1	157	355	69	184	88	171
Initial State		1 1		1 1		0 1	
$\overline{R}$		.233%		.233%		.233%	

	7/62 to 12/87	7/62 to 3/75	4/75 to 12/87
MLE Estimates			
$\hat{\lambda}_{00}$	.576	.622	.511
$\hat{\sigma}(\lambda_{00})$	(.027)	(.035)	(.043)
$\hat{\lambda}_{11}$	.307	.273	.340
$\hat{\sigma}(\lambda_{11})$	(.020)	(.028)	(.029)
Hypothesis Tests			
H_1: $\lambda_{00} = \lambda_{11}$			
LRT	60.1	56.1	10.7
M.S.	(.00)	(.00)	(.00)
H_2: $\lambda_{00} = \lambda_{01} = \lambda_{10} = \lambda_{11}$			
LRT	87.9	85.6	16.0
M.S.	(.00)	(.00)	(.00)

λ_{00} and λ_{11} indicate positive serial correlation in all samples. For the full sample, the LRT of 60.1 for Hypothesis 1 rejects the random walk at the 1 percent marginal significance level. The rejection of the random walk is also evident in the two subsamples with both marginal significance levels less than 1 percent and the level of positive serial correlation greater in the early sample. Furthermore, the data also reject Hypothesis 2, equal transition probabilities, at the 1 percent level.

Test results for weekly returns on a value-weighted portfolio are given in Table VI. Although the point estimates of λ_{00} and λ_{11} indicate positive serial correlation, the LRT rejects Hypothesis 1 only in the early sample, with a marginal significance of 0.04, and not in the full or late sample, with marginal significance of 0.11 and 1.00, respectively. Hypothesis 2 can be rejected for both the full and early sample periods for value-weighted returns,

Table VI

Maximum Likelihood Estimates and Likelihood Ratio Tests for Weekly Nominal Returns on a Value-Weighted Portfolio Based on Fixed Ex Ante Means

The log likelihood function is: $\Sigma_{ij=0}^{1}{}_{0}^{1} N_{ij}\log \lambda_{ij} + M_{ij}\log(1 - \lambda_{ij})$. $\hat{\lambda}_{ij}$ = the estimated transition probability of going from a state sequence of $I_{t-2} = i$ and $I_{t-1} = j$ to $I_t = 0$, where i and j equal 1 for high-return weeks and 0 for low-return weeks, and $\hat{\sigma}(\lambda_{ij})$ = the associated standard deviation. $N_{00} \ldots M_{11}$ = the number of observed occurrences of the associated transitions. The initial state = the values of I_1 and I_2. $\bar{R}$ = the mean weekly return from January 1947 to June 1962. LRT = the likelihood ratio test which is asymptotically distributed χ^2 with 1 and 3 degrees of freedom for H_1 and H_2, respectively, and M.S. = the marginal significance level, which is the probability of obtaining that value of the LRT or higher under the null hypothesis.

Transition Counts		7/62 to 12/87		7/62 to 3/75		4/75 to 12/87	
		0	1	0	1	0	1
	0 0	156	164	93	89	63	75
	0 1	133	171	58	81	74	90
	1 0	164	140	89	51	75	88
	1 1	171	229	82	120	89	109
Initial State		1 1		1 1		0 1	
$\bar{R}$		.241%		.241%		.241%	

	7/62 to 12/87	7/62 to 3/75	4/75 to 12/87
MLE Estimates			
$\hat{\lambda}_{00}$	.487	.511	.457
$\hat{\sigma}(\lambda_{00})$	(.028)	(.037)	(.042)
$\hat{\lambda}_{11}$	.427	.406	.449
$\hat{\sigma}(\lambda_{11})$	(.025)	(.035)	(.035)
Hypothesis Tests			
H_1: $\lambda_{00} = \lambda_{11}$			
LRT	2.6	4.3	0.0
M.S.	(.11)	(.04)	(1.00)
H_2: $\lambda_{00} = \lambda_{01} = \lambda_{10} = \lambda_{11}$			
LRT	10.5	21.0	0.0
M.S.	(.01)	(.00)	(1.00)

although the late sample exhibits remarkably random walk-like behavior. These results agree with the findings of Lo and MacKinlay (1988): 1) the equally-weighted returns show more positive serial correlation than value-weighted returns, and 2) the random walk hypothesis cannot be rejected using the most recent sample of weekly value-weighted returns.[15]

III. Small Sample Statistics and Initial State Information

In this section, the problem of relying on asymptotic tests with few annual observations is addressed with the use of a randomization experiment. Next,

[15]Chow-like LRT tests reject parameter stability across the two periods for the value-weighted weekly returns (marginal significance level of 0.03) but not for the equally-weighted returns. The results of tests for changes in the return generating process are available from the authors.

the consequence of ignoring the initial state information is examined. Finally, an alternative to the asymptotic maximum likelihood tests is considered.

A. Randomization Experiment for Annual Returns

The likelihood ratio test (LRT) is asymptotically distributed as a χ^2 random variable with degrees of freedom equal to the number of restrictions contained in the null hypothesis. Two alternative tests used in maximum likelihood estimations are the Wald test (WT) and the Lagrange multiplier test (LMT) which are also asymptotically χ^2 and thus approximately equal to the LRT in large sample sizes. The choice of test statistics is normally a matter of computational convenience.

The Wald test for Hypothesis 1, $\lambda_{00} = \lambda_{11}$, is:

$$\mathrm{WT} = \frac{\left(\hat{\lambda}_{00} - \hat{\lambda}_{11}\right)^2}{\sigma^2\left(\hat{\lambda}_{11}\right) + \sigma^2\left(\hat{\lambda}_{00}\right)} \sim \chi_1^2 \tag{4}$$

where $\hat{\lambda}_{ii}$ are the *unconstrained* MLE's, and $\sigma^2(\hat{\lambda}_{ii})$ are the standard error estimates, as given by the inverse of the information matrix evaluated using the *unconstrained* MLE's. The Lagrange multiplier test for the same null hypothesis is:

$$\mathrm{LMT} = S^2\left(\tilde{\lambda}_{00}\right) \cdot \sigma^2\left(\tilde{\lambda}_{00}\right) + S^2\left(\tilde{\lambda}_{11}\right) \cdot \sigma^2\left(\tilde{\lambda}_{11}\right) \sim \chi_1^2 \tag{5}$$

where $S(\tilde{\lambda}_{ii})$ are the first partials of the log likelihood function, and $\sigma^2(\tilde{\lambda}_{ii})$ are elements of the inverse of the information matrix, all evaluated using the *constrained* MLE.

When calculated for the annual return portfolios in Section II, the three test statistics are not close to each other, thus raising concerns about the reliability of these asymptotic tests. For example, the real equally-weighted portfolio generates test values of 8.6, 13.0, and 8.1 for the LRT, WT, and LMT, respectively. To overcome this small sample problem, a randomization experiment was developed for the equally-weighted portfolio to determine the model-specific distribution of these statistics given an observation length of 41 with 20 zeros and 21 ones. Similar randomization experiments were conducted based on the observed zero and one counts of the other three portfolios. The null hypothesis requires that returns are serially uncorrelated. The randomization experiment shuffles the sequence of zeros and ones to destroy any time dependence and calculates the test statistics for each reshuffling to estimate its distribution under the null. Ten thousand random sequences of length 41 were generated and the three test statistics were calculated for each sequence. The number of realized test values above a given level, divided by 10,000, was used as an estimate of the significance of that level.[16]

[16] With 10,000 experiments, the marginal significance levels reported in Table II can be shown to have a standard error of less than 0.003. Therefore, only two digits (e.g., 0.05) are used to report marginal significance levels, and 0.00 represents all marginal significance levels below 0.005.

The randomization experiment shows that for Hypothesis 1, all three test statistics overstate the true significance.[17] For example, the benchmark 95 percent level occurs at a LRT value of 4.6 in the real equally-weighted simulation as opposed to 3.8 in a χ_1^2 table. A similar randomization experiment for Hypothesis 2, $\lambda_{00} = \lambda_{01} = \lambda_{10} = \lambda_{11}$, yields the same results, indicating that the χ_3^2 distribution slightly overstates the significance of the LRT. All marginal significance levels reported in the annual tables use the simulation-derived distributions. In contrast, the large number of weekly returns makes the χ^2 asymptotic distribution sufficiently accurate so that simulations are not needed.

B. Incorporation of the Initial State Information

The maximum likelihood estimates in Section II ignore the initial state term π. This term represents the probability of realizing the initial two states and has a specific value of π_{00}, π_{01}, π_{10}, or π_{11}, depending on the actual two-state sequence observed. Billingsly (1961) shows that the effect of the information contained in π on the values of the λ_{ij} parameters will become negligible as the observation length T goes to infinity. For the 1947 to 1987 annual return series T = 41, so the initial state information may be important.

Except for the numerical methods procedure, the incorporation of the initial state information into the estimates of the λ_{ij} generally follows Neftci (1984) with notational changes and with the correction made by Falk (1986, footnote 4).[18] First, the four π_{ab} probabilities are expressed as closed-form functions of the λ_{ij} parameters. Four equations relate the π_{ab} and the λ_{ij} parameters. For example, if the initial two states are $I_1 = 0$ and $I_2 = 0$, the state just preceding the beginning of the sample period has specific probabilities of having a value of $I_0 = 0$ or $I_0 = 1$; $P(I_1 = 0, I_2 = 0) = P(I_2 = 0 \mid I_0 = 0, I_1 = 0) * P(I_0 = 0, I_1 = 0) + P(I_2 = 0 \mid I_0 = 1, I_1 = 0) * P(I_0 = 1, I_1 = 0)$, or $\pi_{00} = \lambda_{00}\pi_{00} + \lambda_{10}\pi_{10}$. Three similar relationships follow and, together with the condition that $\pi_{00} + \pi_{01} + \pi_{10} + \pi_{11} = 1$, form a system with a unique solution for the π_{ab} probabilities in terms of the λ_{ij} parameters. For example, it can be shown that:

$$\pi_{00} = \frac{\lambda_{11}\lambda_{10}}{\lambda_{11}\lambda_{10} + 2\lambda_{11}(1 - \lambda_{00}) + (1 - \lambda_{00})(1 - \lambda_{01})}. \tag{6}$$

Second, parameterized π_{ab} terms are included in the log likelihood function and the resulting MLE's. All four annual portfolios considered over the 1947 to 1987 time period have low initial annual returns, so that $I_1 = 0$ and

[17] The ordering WT > LRT > LMT is shown to be a numerical necessity for *linear* restrictions on the coefficients of certain *linear* models by Berndt and Savin (1977). Breusch (1979) extends this inequality to include *nonlinear* restrictions of *linear* models. Although all of the tests reported in Table II are consistent with this inequality, counter-examples exist which indicate the ordering is not a necessary result for the *nonlinear* Markov chain model.

[18] A brief explanation is also given in Amemiya (1985), Chapter 9.

$I_2 = 0$; consequently, the first partials of the log likelihood function are of the form:

$$\frac{\partial \log L}{\partial \lambda_{ij}} = \frac{\partial \log \pi_{00}}{\partial \lambda_{ij}} + \frac{N_{ij}}{\lambda_{ij}} + \frac{M_{ij}}{1 - \lambda_{ij}} \tag{7}$$

Note that π_{00} is a nonlinear function of all four λ_{ij} parameters, so it is impossible to derive closed-form solutions to the estimates by setting the first partials to zero.

Third, a simple numerical methods approach utilizes an incomplete solution to find MLE's that incorporate the initial states. Substituting the partial derivative of equation (6) into (7), then setting (7) equal to zero and rearranging results in four incomplete MLE solutions of the form:

$$\lambda_{ij} = \frac{N_{ij}}{N_{ij} + M_{ij}} + \frac{\partial \log \pi_{00}}{\partial \lambda_{ij}} \frac{\lambda_{ij}(1 - \lambda_{ij})}{(N_{ij} + M_{ij})} \tag{8}$$

where the partial derivative in the second term can be replaced by its analytical form derived from equation (6). Notice that the first term in equation (8) is exactly the MLE if the initial states are ignored. The relatively small second term includes the asymptotic variance of λ_{ij} and thus decreases in importance as $N_{ij} + M_{ij}$ grows large. The initial estimate of the λ_{ij} parameters are based on the first term in (8); successive estimates of the λ_{ij} parameters are generated by substituting the previous estimate into the right-hand side of (8). Because π_{00} is a function of all four λ_{ij} parameters, the process converges faster when all four λ_{ij} parameters are estimated simultaneously. This functional iterative approach relieves one from analytically deriving the 16 second partials (or 64 if all four π_{ab} probabilities are considered).[19] The analysis and numerical procedure for estimation under the $\lambda_{00} = \lambda_{11}$ constraint follows in a similar manner. With both the unconstrained and constrained MLE's it is possible to calculate the likelihood ratio test of the null hypothesis.

Table VII compares the *simple* MLE and LRT statistics for the real equally-weighted portfolio given in Section II, with the more *precise* initial-state information estimates generated by the iterative procedure. All four portfolios have the same initial-state sequence, $I_1 = 0$ and $I_2 = 0$, which lowers the LRT values. The inclusion of the initial state information has little effect on the estimates of the transition probabilities; for example, $\hat{\lambda}_{00}$ increases by only 0.028, from 0.250 to 0.278 when the initial states are considered. The real equally-weighted LRT falls from 8.6 to 7.8, although the marginal significance level is unchanged. The real value-weighted LRT falls

[19]Neftci (1984) and Falk (1986) employ the Newton-Raphson convergence algorithm that requires second partials. The Newton-Raphson method does exhibit faster second-order convergence speeds than the functional iteration approach used in this section; however, the iteration method converges within 5 iterations and is much easier to derive and program. While the analytically derived second partials can be used to form the information matrix and estimate standard errors, the LRT statistic used in this paper does not require these calculations.

Table VII

A Comparison of Maximum Likelihood Estimates and Likelihood Ratio Tests with and without the Initial State Information for Annual Real Returns on an Equally-Weighted Portfolio (1947-1987)

The log likelihood function is: $L = \log \pi_{00} + \sum_{ij=0\,0}^{1\,1} N_{ij} \log \lambda_{ij} + M_{ij} \log(1 - \lambda_{ij})$. $\hat{\lambda}_{ij}$ = the estimated transition probability of going from a state sequence of $I_{t-2} = i$ and $I_{t-1} = j$ to $I_t = 0$, where i and j equal 1 for high return years and 0 for low return years, and $\hat{\sigma}(\lambda_{ij})$ = the associated standard deviation. $N_{00} \ldots M_{11}$ = the number of observed occurrences of the associated transitions. π_{00} = the probability of initial state values of $I_1 = I_2 = 0$. LRT = the likelihood ratio test, and M.S. = the marginal significance level, which is the probability of obtaining that value of the LRT or higher under the null hypothesis (H_1): $\lambda_{00} = \lambda_{11}$.

	Without Initial State Information	With Initial State Information
Unconstrained MLE		
$\hat{\lambda}_{00}$	.250	.278
$\hat{\lambda}_{01}$	.182	.186
$\hat{\lambda}_{10}$	.500	.537
$\hat{\lambda}_{11}$	.900	.902
L	−19.9	−21.5
Constrained MLE: $\lambda_{00} = \lambda_{11}$		
$\hat{\lambda}_{00}$	.611	.641
$\hat{\lambda}_{01}$	.182	.186
$\hat{\lambda}_{10}$	.500	.532
$\hat{\lambda}_{11}$	.611	.641
L	−24.2	−25.4
LRT	8.6	7.8
M.S.[a]	.01	.01

[a]The M.S. values are from a simulation reported in Section III. *A*.

from 5.8 to 4.9, reducing the significance level by 1 percent to the 4 percent level. The excess equally-weighted and value-weighted portfolios' LRT's fall from 14.7 to 14.1 and from 5.9 to 5.4, but the marginal significance levels are unchanged.[20]

The incorporation of initial-state information slightly reduces the significance of the $\lambda_{00} = \lambda_{11}$ rejection but does not change the conclusion of this paper. The decline in significance is a result of the specific initial states of $I_1 = 0$ and $I_2 = 0$ and is not a general result of incorporating initial state information.

[20]Conducting a randomization investigation of the LRT distribution incorporating initial state information is computationally prohibitive. Each of the 10,000 randomly generated sequences would involve a numerical methods convergence procedure. It is reasonable to assume that the distribution would be only slightly changed; therefore, the noninitial state simulation marginal significance levels are applied to initial state statistics.

C. Application of Fisher's Exact Test

The two-state Markov chain described in Section I reduces annual returns to a binomial series. The null hypothesis is tested by splitting the binomial series into 4 mutually exclusive cases, depending on the previous 2 years. If one is interested in only 2 of the mutually exclusive cases (i.e., Hypothesis 1), then the Fisher (1935) exact probability test is applicable.

The Fisher exact probability test determines whether two groups (prior states of 0 0 and 1 1) differ in the proportion with which they fall into two classifications (above and below $\bar{R}$). In the present context, the advantage of the test is that it gives exact levels of marginal significance, even in small samples, without relying on asymptotic distributions or randomization experiments. The disadvantages are its inability to incorporate initial states and to test higher-order chains or more complex hypotheses (i.e., Hypothesis 2).

By focusing only on Hypothesis 1, the transition count matrix reduces to the following 2×2 contingency table:

	0	1	Total
0 0	N_{00}	M_{00}	$N_{00} + M_{00}$
1 1	N_{11}	M_{11}	$N_{11} + M_{11}$
Total	$N_{00} + N_{11}$	$M_{00} + M_{11}$	T

Fisher shows that when the four marginal totals are regarded as fixed, the probability p of observing a particular set of transition counts in the 2×2 contingency table is given by the hypergeometric distribution:

$$p = \frac{\binom{N_{00} + N_{11}}{N_{00}} \binom{M_{00} + M_{11}}{M_{00}}}{\binom{\mathrm{T}}{N_{00} + M_{00}}} \tag{9}$$

Using equally-weighted annual real returns, the contingency table takes on the following values:

0 0	2	6	8
1 1	9	1	10
Total	11	7	18

The exact probability of observing this set of transition counts under the null hypothesis is found by substitution in equation (9):

$$p = \frac{\binom{11}{2}\binom{7}{6}}{\binom{18}{8}}$$

$$p = 0.00880$$

In addition to the observed transition counts, the following more extreme deviation from the null is possible while conserving the marginal totals:

0 0	1	7	8
1 1	10	0	10
Total	11	7	18

This more extreme case has a probability of 0.00025 under the null.

The probability of observing the historical transaction counts or even more extreme frequencies is 0.00880 + 0.00025 = 0.00905. Since there is no theory to predict the direction of difference (mean reversion or aversion in prices), a two-tailed rejection region must be calculated by doubling the significance level. Thus, Hypothesis 1 can be rejected with a marginal significance level of 0.018. Applying the Fisher exact test to the other 3 annual portfolios yields marginal significance levels of 0.007, 0.063, and 0.057 for the equally-weighted excess, value-weighted real, and value-weighted excess returns, respectively.

A practical difference between the Fisher exact test and the randomization experiments used in Section III.*B* is that the Fisher exact test restricts the marginal totals to equal their historically observed values, whereas randomization allows them to vary. This variation may explain the slight difference in the marginal significance levels reported in Table II and those found using Fisher's exact test and lies at the center of a debate in the statistics literature. On the one hand, considering the marginal totals to be fixed seems arbitrary since they may vary if one could observe repeated samples from the same distributions. On the other hand, the marginal totals are historical outcomes, and performing tests that ignore this information can be thought of as tests on a different set of data. This second line of reasoning requires one to condition on ancillary statistics, such as the marginal totals, when performing tests.[21]

Despite the difference, both the Fisher exact test and the likelihood ratio test used in this paper reject the random walk hypothesis at high levels of significance for postwar annual returns.

IV. Summary

Recent tests based on linear combinations of autocorrelation coefficients by Fama and French (1988) and Poterba and Summers (1988) conclude that stock market prices do not follow a random walk. Specifically, they find mean reverting behavior in prices or negative serial correlation in returns when measured over long horizons. Kim, Nelson, and Startz (1991) show that this mean reverting tendency is overstated due to assumptions of normally dis-

[21]See Cox and Hinkley (1974), pages 31 to 39, for a discussion of ancillary statistics and the conditionality principle.

tributed returns and is a prewar phenomenon only. Our paper introduces a new test of the random walk hypothesis based on the framework of finite-state Markov processes, or Markov chains. The test is nonlinear in the sense that it focuses on periods of established trends and tests whether a low return is more likely after observing a sequence of two high returns or a sequence of two low returns.

Given a time series of returns, a Markov chain is defined by letting state 1 represent high returns and state 0 represent low returns. The random walk hypothesis requires that the probability of state 0 be the same for any prior two-state sequence. In terms of the Markov chain model, this translates into a restriction that the transition probabilities are equal.

Annual real and excess returns are shown to exhibit significant nonrandom walk tendencies in the sense that low (high) returns tend to follow runs of high (low) returns. The random walk is rejected at the 1 percent marginal significance level using postwar calendar year real returns for the equally-weighted portfolio and at the 5 percent marginal significance level for the value-weighted portfolio. The marginal significance levels are calculated using a randomization test and with Fisher's exact test since asymptotic tests overstate the rejection of the random walk. This rejection of the random walk is for the *postwar* period, unlike the rejections of Fama and French (1988) and Poterba and Summers (1988). The model does not reject the random walk at conventional levels of significance for the prewar period. The rejection of the random walk weakens as annual returns are calculated using starting months other than January.

The Markov chain tests also confirm the findings of Lo and MacKinlay (1988) with below (above) average weekly returns often following below (above) average returns. Small stocks are shown to exhibit more positive weekly serial correlation than large stocks, and large stocks are shown to follow a random walk over the last decade.

REFERENCES

Amemiya, T., 1985, *Advanced Econometrics*, Chap. 9, (Harvard University Press, Cambridge, MA).

Anderson, T. W., 1984, *An Introduction to Multivariate Statistical Analysis*, 2nd Edition, Chap. 4, (John Wiley & Sons, Somerset, NJ).

Berndt, E. R. and N. E. Savin, 1977, Conflict among criteria for testing hypotheses in the multivariate regression model, *Econometrica* 45, 1263-1278.

Billingsley, P., 1961, *Statistical Inference for Markov Processes*, (Chicago Press, Chicago, IL).

Blanchard, O. S. and M. Watson, 1982, Bubbles, rational expectations and financial markets, in P. Wachtel, ed.: *Crises in the Economic and Financial Structure*, (Lexington Books, Lexington, MA).

Breusch, T. S., 1979, Conflict among criteria for testing hypotheses: Extensions and comments, *Econometrica* 47, 203-206.

Cecchetti, S. G., P. Lam, and N. C. Mark, 1990, Mean reversion in equilibrium asset prices, *The American Economic Review* 80, 398-418.

Chang, E. C. and J. M. Pinegar, 1991, The predictive power of January returns in foreign and domestic markets, *Economics Letters*, Forthcoming.

Cox, D. R. and D. V. Hinkley, 1974, *Theoretical Statistics*, (Chapman and Hall, London).

Dryden, M. M., 1969, Share price movements: A Markovian approach, *Journal of Finance* 24, 49-60.

Engel, C. and J. D. Hamilton, 1990, Long swings in the exchange rate: Are they in the data and do markets know it?, *The American Economic Review* 80, 689-713.

Evans, G. W., 1986, A test for speculative bubbles in the sterling dollar exchange rate: 1981-1984, *The American Economic Review* 76, 621-636.

Falk, B., 1986, Further evidence on the asymmetric behavior of economic time series over the business cycle, *Journal of Political Economy* 94, 1096-1109.

Fama, E., 1970, Efficient capital markets: A review of theory and empirical work, *The Journal of Finance* 25, 383-416.

—— and K. French, 1988, Permanent and temporary components of stock prices, *Journal of Political Economy* 96, 246-273.

Fielitz, B. D., 1975, On the stationarity of transition probability matrices of common stocks, *Journal of Financial and Quantitative Analysis* 10, 327-339.

—— and T. N. Bhargava, 1973, The behavior of stock-price relatives: A Markovian analysis, *Operations Research* 21, 1183-1199.

Fisher, R. A., 1935, *Statistical Methods For Research Workers*, 5th Edition, (Oliver and Boyd, Edinburgh).

Gregory, A. W. and M. Sampson, 1987, Testing the independence of forecast errors in the forward foreign exchange market using Markov chains: A cross-country comparison, *International Journal of Forecasting* 3, 97-113.

Hamilton, J. D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* 57, 357-384.

Ibbotson, R. G. and R. A. Sinquefield, 1988, *Stocks, Bonds, Bills, and Inflation 1988 Yearbook*, (Ibbotson Associates, Chicago, IL).

Keim, D. B., 1983, Size related anomalies and stock return seasonality: Further empirical evidence, *Journal of Financial Economics* 12, 13-32.

Kim, M., C. Nelson, and R. Startz, 1991, Mean reversion in stock prices? A reappraisal of the empirical evidence, *Review of Economic Studies*, Forthcoming.

Kroll, Y., H. Levy, and A. Rapoport, 1988, Experimental tests of the mean-variance model of portfolio selection, *Organizational Behavior and Human Decision Process* 42, 388-410.

Ljung, G. M. and G. E. P. Box, 1978, On a measure of lack of fit in time series models, *Biometrika* 65, 297-303.

Lo, A. and A. MacKinlay, 1988, Stock market prices do not follow random walks: Evidence from a simple specification test, *The Review of Financial Studies* 1, 41-66.

Neftci, S. N., 1984, Are economic time series asymmetric over the business cycle?, *Journal of Political Economy* 92, 307-328.

Niederhoffer, V. and M. F. M. Osborne, 1966, Market making and reversal on the stock exchange, *Journal of the American Statistical Association* 61, 897-916.

Poterba, J. and L. Summers, 1987, *Mean Reversion in Stock Prices: Evidence and Implications*, (Mimeograph, NBER, Cambridge, MA).

—— and L. Summers, 1988, Mean reversion in stock prices: Evidence and implications, *Journal of Financial Economics* 22, 27-59.

Reinganum, M. R., 1983, The anomalous stock market behavior of small firms in January, *Journal of Financial Economics* 12, 89-104.

Rozeff, M. S. and W. R. Kinney, 1976, Capital market seasonality: The case of stock returns, *Journal of Financial Economics* 3, 379-402.

Samuelson, P., 1988, Longrun risk tolerance when equity returns are mean regressing: Pseudoparadoxes and vindication of business man's risk, presented at the Colloquium for James Tobin, Yale University.

Turner, C. M., R. Startz, and C. R. Nelson, 1989, A Markov model of heteroskedasticity, risk, and learning in the stock market, *Journal of Financial Economics*, 25, 3-22.