Review of Economic Studies (1991) **58**, 515-528 0034-6527/91/00310515$02.00

Mean Reversion in Stock Prices? A Reappraisal of the Empirical Evidence

MYUNG JIG KIM
University of Alabama

CHARLES R. NELSON
University of Washington and NBER

and

RICHARD STARTZ

University of Washington

First version received September 1989; *final version accepted August* 1990 (*Eds.*)

The paper re-examines the empirical evidence for mean-reverting behaviour in stock prices. Comparison of data before and after World War II shows that mean reversion is entirely a pre-war phenomenon. Using randomization methods to calculate significance levels, we find that the full sample evidence for mean reversion is weaker than previously indicated by Monte Carlo methods under a Normal assumption. Further, the switch to mean-averting behaviour after the war is about to be too strong to be compatible with sampling variation. We interpret these findings as evidence of a fundamental change in the stock returns process and conjecture that it may be due to the resolution of the uncertainties of the 1930's and 1940's.

1. INTRODUCTION

A stylized version of efficient markets theory states that the sequence of holding period returns on a risky asset should be serially random. A large body of empirical literature seemed to support the theory for stock prices (see Fama (1970) and LeRoy (1982) for surveys and discussion), finding no evidence of serial correlation. However, a recent series of papers including those by Poterba and Summers (1988) (herafter P&S), Lo and MacKinlay (1988), and Clark (1987) challenge this conventional view, using the variance-ratio methodology of Cochrane (1988). The variance ratio at lag K is defined as the ratio of the variance of the K-period return to the variance of the one-period return divided by K, which is unity under the random-walk hypothesis. The empirical evidence presented in these studies shows that sample variance ratios are typically below unity for lags longer than a year and above unity for shorter lags. The variance ratio can be thought of as summarizing the autocorrelations of returns since the sample variance ratio can be expressed as a positively weighted sum of the sample autocorrelations plus unity. Thus, stock returns seem to be characterized by positive autocorrelation over intervals under a year and by negative autocorrelation over longer intervals. The latter finding has been interpreted as evidence of "mean reverting" behaviour in stock prices. Evidently, a given change in price tends to be reversed over the next several years by a predictable change in the opposite direction. If true, this would suggest that there are transitory deviations from equilibrium which are both large and persistent. Using a closely related methodology based on autoregressions of multilayer returns, Fama and French (1988) (hereafter F&F) also found evidence of mean reversion and conclude that about 40% of the variation in stock returns is predictable from past returns.

This paper re-examines the finding of mean-reverting behaviour in stock prices presented by the recent literature. Comparison of the historical evidence for sample periods before and after World War II suggests that mean reversion was a pre-war phenomenon. Post-war variance ratios do not in general display evidence of mean reversion, but rather suggest persistence in returns which we will refer to as "mean aversion". It is not surprising therefore that the autoregressions fitted by Fama and French have no predictive power after World War II. Refitting them over the post-war period results is a reversal of the sign of the coefficient from negative to positive, again suggesting mean aversion rather than reversion. Measures of statistical significance reported in the literature have been based on Monte Carlo simulations assuming Normal disturbances. Actual stock returns are generally recognized to be non-Normal, leading to possibly incorrect inferences. This paper presents estimates of the unknown distributions of the variance ratio and F&F autoregression statistics using randomization methods. The results suggest that significance levels are much lower than previously reported. They also suggest that evidence of mean aversion for the post-war period is roughly as strong as the evidence of mean reversion for the whole period. Further, the observed differences between pre- and post-war variance ratios and autoregression coefficients are shown to be too large to be consistent with a regime of random returns over the whole period. The evidence suggests a change in the structure of returns at the end of World War II.

This paper is organized as follows. Section 2 briefly describes the background of this study and examines the historical variance ratios for the monthly return series from the CRSP file for 1926–86. Excess returns above the short-term interest rate and real returns deflated by the consumer price index (CPI) are considered. We discuss the randomization method for approximating the unknown sampling distribution of the test statistic and its exact significance level. We also investigate the effect of the large variance of stock market returns in 1930s by introducing the idea of stratified randomization. Section 3 considers the alternative but closely related test for mean-reversion proposed by Fama and French (1988) based on predictability of multiperiod returns. We find that their apparent finding that returns are negatively related to past returns is also a pre-war phenomenon. Section 4 concludes the paper.

2. VARIANCE RATIOS BEFORE AND AFTER WORLD WAR II

The variance-ratio test is motivated by the notion that if the underlying data generating process for stock returns is serially random with constant variance, then the variance of the return over K periods is simply $K\sigma^2$, where σ^2 is the variance of the one-period return. Therefore, the simplest version of variance-ratio test calculates the statistic

$$VR(K)=\frac{\text{Var}\,(r_t^K)}{\text{Var}\,(r_t^1)}\cdot\frac{1}{K} \tag{1}$$

where r_t^K is the K-period return. The null hypothesis of a random walk is rejected if this statistic is significantly different from unity. Cochrane (1988) showed that $VR(K)$ can be approximated by

$$VR(K)\approx 1+2\textstyle\sum_{j=1}^{K-1}\frac{(K-j)}{K}\hat{\rho}(j) \tag{2}$$

where $\hat{\rho}(j)$ denotes the j-th-order sample autocorrelation coefficient of the one-period stock return. Equation (2) makes clear the relation between the sample autocorrelations of one-period returns and the variance ratio. The expected value of $VR(K)$ under the null hypothesis of serial independence of returns is derived by noting that the j-th-order sample autocorrelation has expected value $-1/(T-j)$ as shown in Kendall and Stuart (1976), so that

$$E[VR(K)]=\frac{2-K}{K}+\frac{2}{K}\sum_{j=1}^{K-1}\frac{T-K}{T-j}. \tag{3}$$

Dividing by this quantity provides a bias correction for the sample variance ratio. For monthly returns series, the variance ratio is expressed in terms of the variance of returns over integer multiples of twelve months relative to the variation over a one-year span as in P&S. Letting K denote years and $k=12K$ the number of months, the formula becomes

$$VR(K)=\frac{\mathrm{Var}\,(r_t^k)/K}{\mathrm{Var}\,(r_t^{12})/12} \tag{4}$$

$$\approx 1+2\sum_{j=1}^{K-1}\left(\frac{K-j}{K}\right)\hat{\rho}(j)-2\sum_{j=1}^{11}\left(\frac{12-j}{12}\right)\hat{\rho}(j) \tag{5}$$

where $r_t^K=\sum_{i=0}^{k-1} r_{t-i}$ and $\hat{\rho}(j)$ is the j-th sample autocorrelation of monthly returns. P&S derive the expected value of $VR(K)$ which provides a bias correction for monthly data which we incorporate in our calculations.

The data set consists of monthly total returns on all NYSE stocks from the CRSP files for both value-weighted (hereafter VW) and equal-weighted (hereafter EW) portfolios from 1926 through 1986. The one-month Treasury-bill rate and the CPI for all urban consumers (not seasonally adjusted) from Ibbotson Associates are used to calculate excess and real returns respectively. Since P&S report significance levels only for excess returns, we report in detail only the results for excess returns, but indicate generally how those for real returns differ. Complete tabulations are available from the authors.

In Figure 1(a) are plotted the sample variance ratios for excess returns on the VW and EW portfolios using the P&S formula for monthly data with $K=1,\ldots,10$ years. Results for the full sample period 1926–1986 correspond closely to those reported by P&S for 1926–1985 and K up to 8 years. (P&S also discussed one-month returns which we do not deal with in this paper.) As noted by P&S, the VR for the full sample period generally declines with increasing lag and is well below unity at lags of several years. However, extending K to 10 years does not result in smaller VRs. Further, mean reversion

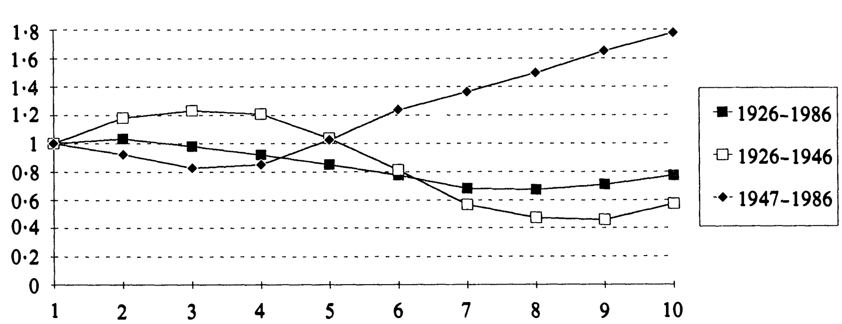

FIGURE 1(a)
Variance ratios for excess return on CRSP value-weighted portfolio

is stronger for the EW portfolio than the VW. P&S estimated significance levels for *VR* using Monte Carlo methods under the null hypothesis that returns are random. In the case of *VR*(8) for excess returns they reported *p*-values of 0·08 for VW and 0·005 for EW.

Now consider the sample estimates of *VR* for the subperiods defined by the end of World War II which are also plotted in Figure 1(*b*). The sample period 1926–1946 displays an even more severe decline at longer lags. Minimum values occur at $K = 9$ years. However, for the post-war period the story is very different. In the case of the VW portfolio, the *VR*s drop below 1 at low lags but then rise with lag until at lag 10 years they are well above one. For the EW portfolios the *VR*s form a U-shaped pattern, declining to about 0·7 at lag 4 years but then rising to around one at lag 10.

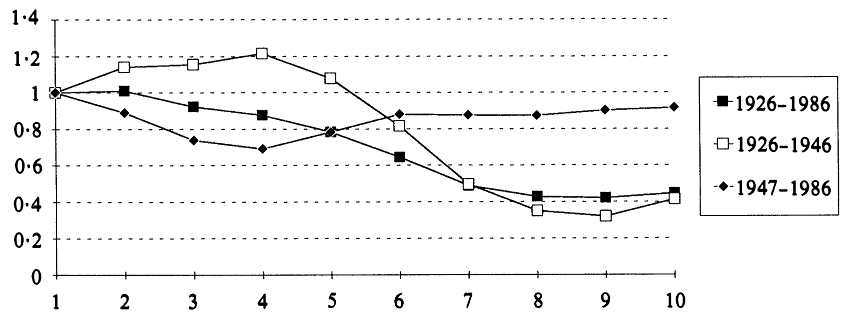

FIGURE 1(*b*)
Variance ratios for excess return on CRSP equal-weighted portfolio

P&S reported partial results for the subsample 1936–1985 which excludes the first 10 years of the CRSP data. They found that results for the EW portfolio were robust to this change, but that for the VW portfolio *VR* rose to about unity at lag 8 years. Taking a traditional and what seems to us a more economically appropriate dividing point at the end of World War II, we find a sharp difference between the earlier and later periods. While mean reversion only disappears for EW returns, it switches sign in the case of VW returns. The contrasts between sub-periods were even stronger for real returns. This suggests that if sample values are indicative of the underlying stochastic process then the post-war period was characterized by mean aversion, a tendency towards persistence rather than reversal in stock price movements. We next consider whether deviations of *VR* from unity are significant according to randomization methods of estimating the sampling distribution, and whether they are significantly different before and after World War II.

The small sample distribution of *VR* has been derived analytically only for the Normal case by Faust (1989). Lo and MacKinlay (1988, 1989) have treated the *VR* as a Hausman specification test to obtain asymptotic standard errors and have studied the power and size of the *VR* test using Monte Carlo methods under Normality with heteroscedasticity. P&S and F&F also include with their results standard errors based on Monte Carlo methods. Recently, Richardson and Stock (1990) have developed approximations based on an asymptotic theory which treats the degree of overlap in the data, K/T, as tending to a fixed fraction. That stock returns are non-Normal and heteroscedastic is well-known. The actual distribution is of course unknown. The bootstrap and randomization (or shuffling) methods are appropriate when the population distribution is unknown, since they both rely on re-sampling the data to estimate the distribution of sample statistics. The question addressed by the bootstrap method is, as is stated in Efron (1979), given a random sample $x=(x_1, x_2, \ldots, x_T)$ from an unknown probability distribution F, estimate the sampling distribution of some prespecified random variable $R(x, F)$, on the basis of the observed data x. Namely, we proceed as if the sample is the population for purposes of estimating the sampling distribution of the test statistic, $R(x, F)$. Randomization differs from bootstrapping in that it addresses the question of whether or not there is a relationship between variables, regardless of the nature of stochastic distrubances.

The idea of randomization tests appeared in the literature as early as Fisher (1935). Noreen (1989) provides a clear and practical exposition of Monte Carlo, bootstrap and randomization methods. Randomization focuses on the null hypothesis that one variable is distributed independently of another. In our context the null hypothesis we are interested in is that returns are distributed independently of their ordering in time. Randomization shuffles the data to destroy any time dependence and then recalculates the test statistic for each reshuffling to estimate its distribution under the null. Repeating the experiment, we count how many times the calculated variance ratio after randomization falls below the value of the actual historical statistic to estimate the significance level. The advantage of this approach is that the null hypothesis is very simple and no assumptions are made concerning the distribution of stock prices. Furthermore, as Noreen points out, the data do not have to be a random sample from some specific population distribution function. The practical difference between randomization and bootstrap is that in the latter method we sample the data with replacement. In each case our estimates are based on 1000 shuffles.

Randomization estimates of the moments and fractiles of the *VR* for the VW and EW portfolios as well as the implied significance levels for historical sample *VR*s are presented in Table I. We note that the means of the sample *VR*s are close to unity, suggesting that the P&S bias correction works well. The estimated standard deviation, SD, is compared with the standard deviation implied by relation (5) between *VR* and $\hat{\rho}$, using the approximation for the sample variance of $\hat{\rho}$ due to Bartlett (1946), denoted SD^B. The Bartlett approximation gives too large a standard deviation at lags 2 and 3, but much too small a standard deviation at long lags. It is important to calculate p-values exactly rather than using a standard deviation under an assumption of Normality for *VR* because its sampling distribution is skewed. The randomization method implies substantially weaker significance than reported by P&S. Instead of a p-value of 0·08 at lag 8 years for VW returns we estimate 0·197, and the smallest is at lag 7 years with 0·183. Similarly, the estimated p-value for EW excess returns at lag 8 years is larger than reported by P&S, namely 0·03 compared with 0·005. We note that evidence of mean reversion is stronger for EW than for VW returns, and stronger for real than for excess returns.

A number of empirical studies such as Officer (1973) and more recently Schwert (1989) have documented evidence that the variance of stock prices was much higher during the period 1929–1940 than before or since. Studying market factor variability over the period 1897–1969 Schwert finds that variability returned to pre-Depression levels after the 1930s. To see whether a change in the variance of returns might affect the sampling distribution of *VR* we have carried out a stratified randomization of the data. In particular, the returns from the high-variance period 1930–1939 are placed in a separate urn from which returns for those months are drawn without replacement when generating artificial histories. The results reported in Table II suggest somewhat stronger significance for VW returns (p-value of 0·146 at lag 8 years) but weaker results for EW returns (p-value of

TABLE I

Randomization estimates of the sampling distribution of the sample VR; CRSP *monthly excess returns* 1926–1986

K	2	3	4	5	6	7	8	9	10
				Value-weighted portfolio					
Mean	1·000	1·001	1·001	1·001	1·001	1·003	1·005	1·008	1·013
Median	0·999	0·994	0·986	0·974	0·977	0·950	0·944	0·936	0·918
SD	0·105	0·175	0·229	0·275	0·319	0·359	0·397	0·436	0·472
SD^B	0·216	0·251	0·272	0·281	0·281	0·267	0·281	0·314	0·360
				Fractiles					
5%	0·827	0·717	0·658	0·601	0·544	0·497	0·461	0·435	0·409
10%	0·860	0·771	0·716	0·658	0·632	0·581	0·550	0·525	0·502
20%	0·915	0·856	0·800	0·764	0·724	0·695	0·673	0·635	0·617
Sample	1·035	0·980	0·919	0·849	0·775	0·682	0·671	0·709	0·771
p-value	0·642	0·456	0·375	0·307	0·252	0·183	0·197	0·269	0·345
				Equal-weighted portfolio					
Mean	0·999	1·000	0·999	1·000	0·999	0·999	1·000	1·003	1·006
Median	0·997	0·995	0·986	0·963	0·960	0·946	0·933	0·913	0·905
SD	0·107	0·176	0·229	0·275	0·319	0·360	0·339	0·438	0·475
SD^B	0·211	0.236	0·259	0·259	0·234	0·190	0·179	0·187	0·208
				Fractiles					
5%	0·820	0·724	0·649	0·600	0·559	0·517	0·482	0·444	0·414
10%	0·860	0·775	0·714	0·675	0·630	0·580	0·548	0·510	0·492
20%	0·910	0·845	0·793	0·758	0·712	0·693	0·665	0·638	0·621
Sample	1·009	0·923	0·877	0·783	0·646	0·487	0·427	0·421	0·445
p-value	0·538	0·347	0·322	0·237	0·117	0·040	0·030	0·038	0·070

0·059 at lag 8 years). Stratification also weakens the significance of *VR*s based on real returns. The high-volatility period differs in mean as well as dispersion. To see how important is the difference in mean return we did an additional stratified randomization in which sub-period means are subtracted out of each observation and replaced by the mean of the whole sample. With the mean return for each sub-period thereby equalized, the estimated significance of the *VR*s is weaker than reported here.

Statements about the significance of *VR*s have so far focussed on the smallest *p*-values across all lags that were included. Of course, the probability of obtaining a *VR* relatively far from 1 at *some* lag is higher than at a predetermined lag. We and others may therefore have overstated the significance of the results; see Richardson and Stock (1990) for further discussion. Ideally we would like to be able to calculate the significance of the *VR*s across all lags. This would involve either a parametric approximation to that distribution which we lack or an extraordinary amount of computing which we find impractical. To try to get at the same objective we have computed the probability of getting at least one *VR* among the nine that is individually significant at a given level under stratified randomization. The lowest *p*-value for VW excess returns was 0·135, occurring at lag 7, but the experiment shows that the probability of finding at least one *VR* at that level or better is 0·287. Similarly, the best significance level for EW excess returns at an individual lag is 0·059, but the probability of finding at least one at that level is 0·162.

TABLE II

Stratified randomization estimates of the sampling distribution of the sample VR; CRSP *monthly excess returns* 1926–1986

K	2	3	4	5	6	7	8	9	10
					Value-weighted portfolio				
Mean	1·007	1·021	1·031	1·031	1·028	1·023	1·018	1·013	1·008
Median	1·003	0·998	0·999	0·981	0·977	0·974	0·980	0·975	0·960
SD	0·128	0·200	0·250	0·287	0·313	0·332	0·344	0·353	0·360
					Fractiles				
5%	0·800	0·731	0·687	0·637	0·603	0·567	0·527	0·501	0·501
10%	0·845	0·783	0·744	0·697	0·670	0·649	0·626	0·598	0·575
20%	0·894	0·845	0·814	0·785	0·759	0·738	0·722	0·718	0·717
Sample	1·035	0·980	0·919	0·849	0·775	0·682	0·671	0·709	0·771
p-value	0·598	0·468	0·371	0·286	0·222	0·135	0·146	0·189	0·258
					Equal-weighted portfolio				
Mean	0·974	0·960	0·941	0·914	0·883	0·852	0·822	0·794	0·769
Median	0·971	0·939	0·900	0·860	0·817	0·793	0·763	0·746	0·718
SD	0·141	0·217	0·267	0·298	0·316	0·323	0·323	0·321	0·317
					Fractiles				
5%	0·746	0·642	0·582	0·515	0·465	0·426	0·394	0·366	0·354
10%	0·788	0·693	0·637	0·579	0·523	0·494	0·464	0·426	0·407
20%	0·846	0·766	0·702	0·655	0·613	0·578	0·549	0·527	0·503
Sample	1·009	0·923	0·877	0·783	0·646	0·487	0·427	0·421	0·445
p-value	0·600	0·480	0·452	0·386	0·238	0·089	0·059	0·094	0·139

We now address the question of whether the difference in variance ratios pre- and post-World War II is simply within the range of sampling variation, or whether it suggests that a structural shift occurred between the two periods. *VR*s for VW returns rise to nearly 1·8 at lag 10 years for the post-war sample period 1947–1986. Does this constitute significant evidence of mean *aversion* in the post-war period? Is the difference between sample periods too large? We have computed the randomization significance levels of post-war *VR*s for VW returns where now the relevant *p*-value is the probability in the upper tail. The *p*-value at lag 10 is 0·106. Evidently for VW returns there is as much evidence of mean aversion for the post-war period as there is of mean reversion for the whole period. We also estimated the probability that the pre- and post-war values of *VR* would be as different as observed if returns were random. The *p*-values at lags 9 and 10 years where the contrast is greatest are 0·068 and 0·072 respectively. Real returns provide corresponding results of 0·041 and 0·083. We take these *p*-values as evidence that the end of World War II may have marked a change in the generating process for stock returns to one in which mean reversion is absent.

3. EVIDENCE BASED ON AUTOREGRESSIONS OF MULTIPERIOD RETURNS

A close relative of variance ratios is the regression coefficient calculated by Fama and French (1988, 1989) and Fama (1990). They regress the cumulative return from period

t to period $t+K$ on the return from $t-K$ to t, so their estimating equation is

$$r_{K,t+K}=\alpha_K+\beta_K r_{K,t}+\varepsilon_{K,t+K}. \tag{6}$$

The OLS estimate of β_K is closely related to the VR since

$$\hat{\beta}_K=\frac{\hat{\rho}_1+2\hat{\rho}_2+\cdots+K\hat{\rho}_K+(K+1)\hat{\rho}_{K+1}+\cdots+\hat{\rho}_{2K-1}}{K+2[(K-1)\hat{\rho}_1+\cdots+\hat{\rho}_{K-1}]}. \tag{7}$$

Thus, both VR and β are functions of sample autocorrelations of the consecutive one-month return series. Both exploit the same sample information, but with different weights. While VR is distributed around one for a random walk, β is distributed around zero with negative values indicating mean reversion.

The OLS estimated β are plotted in Figure 2(a) for real returns in order to be comparable to F&F. We will comment on results for excess returns, and full details are available from the authors. For the full sample period, evidence for mean reversion comes from the negative values of β at lags 3, 4 and 5 years particularly. F&F follow Hansen and Hodrick (1980) in calculating standard errors that adjust for the positive autocorrelation in residuals that is induced by overlapping observations. Before computing t-ratios, F&F also adjust for downward bias in the OLS β which they estimate by Monte Carlo. They report that the null hypothesis that returns are random can be rejected in favour of the alternative of mean reversion at roughly the 0·05 level in the case of VW real returns and the 0·002 level in the case of EW real returns, on the basis of the full sample. Figure 2(b) shows the β estimates for the 1926–1946 and 1947–1986 sub-periods.

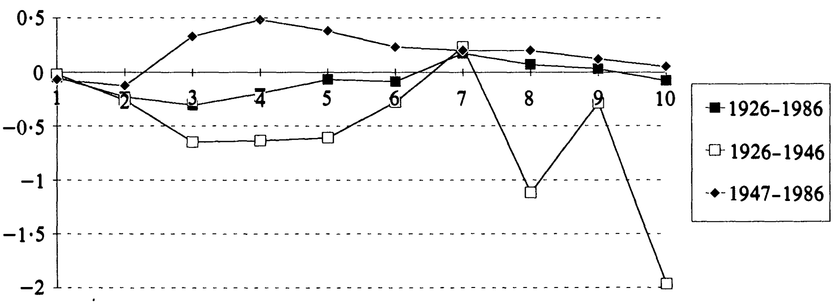

FIGURE 2(a)

AR slope coefficient for real return on CRSP value-weighted portfolio

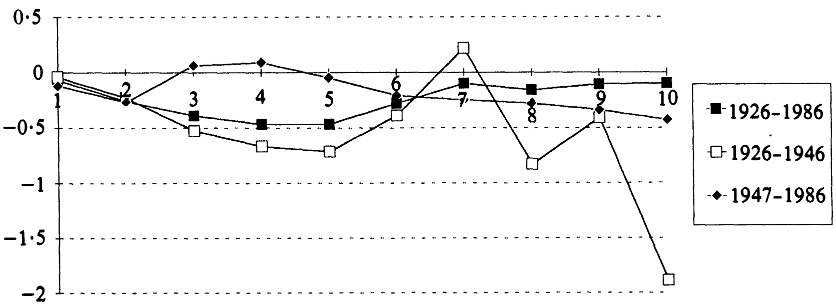

FIGURE 2(b)

AR slope coefficient for real return on CRSP equal-weighted portfolio

As in the case of the *VR*, it is the pre-1947 period that accounts for the evidence of mean reversion. While β takes a large negative value in the 3 to 5 year lag range for the 1926–1946 data, it is positive for the post-war VW returns and around zero for post-war EW returns. A similar pattern holds for excess returns. Indeed, F&F report that bias-adjusted β is around zero for the 1941–1985 subperiod, but they conclude that reliable contrasts across sub-periods are impossible in view of large standard errors.

The randomization technique is used here to investigate the sampling distribution of the β statistic. The null hypothesis to be tested is again that returns are drawn independently regardless of the underlying distribution. We first run the regression using OLS. Second, we randomize the original return series and construct the new K-year holding period returns and run OLS again. Third, to estimate the sampling distribution of coefficients we repeat the procedure 1000 times. The results for real returns 1926–1986 are presented in Table III. As reported by F&F, the OLS estimate of β is biased downward, and our estimates of the bias are in close agreement with theirs. Randomization estimates of the standard deviations, however, are larger than H&H estimates. Reflecting this, estimated significance levels are weaker. The p-value for VW returns at lag 3 years is 0·094. Similar significance levels are implied by a t-statistic calculated using the randomization estimates of bias and standard deviation. In contrast, the p-value implied by the H&H standard error is 0·055. For EW real returns we obtain a best p-value of 0·028 at lag 4 instead of the much smaller 0·002 p-value implied by the H&H standard error. We note here that the obvious advantage of the randomization method in this case is simplicity. All we need to do to estimate significance is to run OLS, not worrying about small sample properties of estimated standard deviations. The asymptotic approximation of H&H evidently produces ones which are too small.

Stratified randomization is again used to see if the apparent change in variance in the 1929–1941 period affects our inferences. The significance levels for VW portfolios are not greatly changed, but for EW returns they weaken further, in particular to 0·062 at lag 4.

Note again that evidence for mean reversion is apparent only at specific lags. Ten out of the 20 estimated βs reported in Table III are in the upper half of the estimated distribution. If we have no prior basis for choosing a particular lag we may be overstating significance by focussing on the lag with the lowest p-value. For that matter, one can make a case for mean aversion at long lags. Following the procedure described for *VR*s, we calculate the probability of obtaining at least one β with a p-value at or below a given level at some lag. For VW real returns the combined significance is 0·375. For EW real returns the combined significance level is 0·339. Similar results are obtained for excess returns.

The shift of the sign of β from negative in the pre-war period to positive in the post-war period for VW returns raises the question of whether the latter may have been a period of significant mean aversion in stock prices. We found that upper-tail significance levels implied by randomization are very high. The probability that β would be larger than that observed at lag 4 is less than 0·01 for both excess and real returns in the post-war data. The evidence for mean aversion in the post-war period is at least as persuasive as the evidence for mean reversion over the whole period. While mean reversion may have been a feature of the pre-war environment, it has not been since.

As in the case of *VR*s we would like to know whether the difference between pre- and post-war βs is too large, suggesting a change in the generating process. We randomized the monthly returns, calculating βs for pre- and post-war periods and tabulating the probability of obtaining a difference as large as that in the historical data. For VW returns

TABLE III

Randomization estimates of the sampling distribution of the OLS *slope* β_K; CRSP *monthly real returns* 1926–1986

Value-weighted portfolio

K	1	2	3	4	5	6	7	8	9	10
Mean	−0·02	−0·04	−0·07	−0·10	−0·12	−0·15	−0·18	−0·21	−0·24	−0·26
Median	−0·02	−0·04	−0·07	−0·10	−0·12	−0·15	−0·18	−0·22	−0·24	−0·28
SD	0·10	0·15	0·18	0·20	0·22	0·24	0·26	0·28	0·29	0·31
					Fractiles					
5%	−0·19	−0·29	−0·36	−0·42	−0·48	−0·55	0·25	0·26	0·26	−0·73
10%	−0·16	−0·24	−0·31	−0·36	−0·42	−0·47	0·18	0·16	0·14	−0·65
20%	−0·11	−0·17	−0·22	−0·27	−0·32	−0·36	0·05	0·03	0·01	−0·52
Sample	−0·05	−0·23	−0·31	−0·20	−0·07	−0·09	0·17	0·07	0·03	−0·08
p-value	0·389	0·113	0·094	0·326	0·588	0·604	0·896	0·835	0·813	0·738
				t-ratio (in parenthesis) and one-sided significance level						
From randomization	(−0·30)	(−1·27)	(−1·30)	(−0·50)	(0·23)	(0·25)	—	—	—	—
	0·382	0·102	0·097	0·309	0·409	0·401	—	—	—	—
implied by H&H	(−0·27)	(−1·36)	(−1·60)	(−0·63)	(0·30)	(0·30)	—	—	—	—
	0·394	0·087	0·055	0·264	0·382	0·382	—	—	—	—
					Equal-weighted portfolio					
Mean	−0·02	−0·04	−0·06	−0·09	−0·12	−0·15	−0·17	−0·20	−0·23	−0·27
Median	−0·02	−0·03	−0·08	−0·10	−0·13	−0·15	−0·17	−0·21	−0·25	−0·28
SD	0·10	0·15	0·18	0·21	0·23	0·24	0·26	0·28	0·30	0·31
					Fractiles					
5%	−0·19	−0·29	−0·37	−0·42	−0·48	−0·53	−0·60	−0·66	−0·72	−0·75
10%	−0·16	−0·23	−0·30	−0·36	−0·41	−0·47	−0·52	−0·57	−0·61	−0·65
20%	−0·11	−0·17	−0·22	−0·27	−0·32	−0·35	−0·41	−0·45	−0·49	−0·52
Sample	−0·07	−0·26	−0·39	−0·47	−0·47	−0·28	−0·10	−0·16	−0·11	−0·10
p-value	0·307	0·069	0·039	0·028	0·059	0·300	0·603	0·561	0·667	0·707
				t-ratio (in parenthesis) and one-sided significance level						
From randomization	(−0·50)	(−1·47)	(−1·83)	(−1·81)	(−1·52)	(−0·54)	—	—	—	—
	0·309	0·071	0·034	0·035	0·064	0·295	—	—	—	—
implied by H&H	(−0·45)	(−1·57)	(−2·36)	(−2·85)	(−2·43)	(−0·71)	—	—	—	—
	0·326	0·058	0·009	0·002	0·007	0·239	—	—	—	—

the p-value is below 0·02 at lags 3, 4 and 10. For EW returns the differences are smaller and less significant. We take these results to suggest a change in the stochastic structure of returns after World War II.

The F&F approach lends itself to post-sample testing since it provides a predictive relation for future returns. Given an estimate of β and the intercept, the previous multi-year return implies the predicted value of the next. In the first experiment we use the OLS values estimated for 1926–1946 to forecast from each successive month the real return on the value-weighted portfolio for the next three years. The choice of horizon is motivated by the fact that the results appear most promising for $K = 3$ years. The resulting correlation between actuals and predictions for the post-war period is $-0{\cdot}08$. We also used the sample estimates based on the whole period 1926–1986, those which gave the best fit after the fact. The correlation is again $-0{\cdot}08$. To simulate real-time forecasting we also forecasted ahead from each month based on coefficients estimated up through that month and the results are plotted in Figures 3(a) and 3(b) for both VW and EW

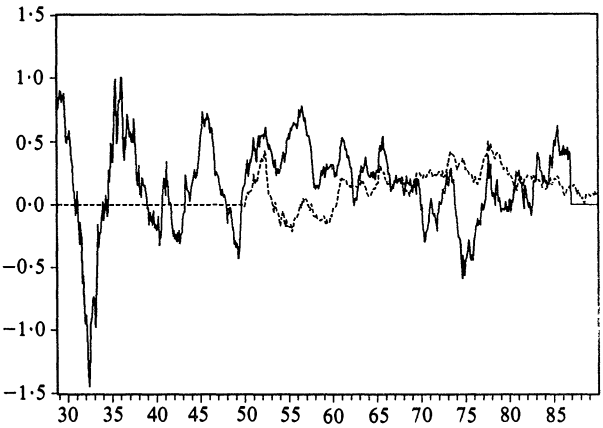

FIGURE 3(a)
3-year ahead predictions of 3-year VW returns: based on the recursive OLS updating procedure

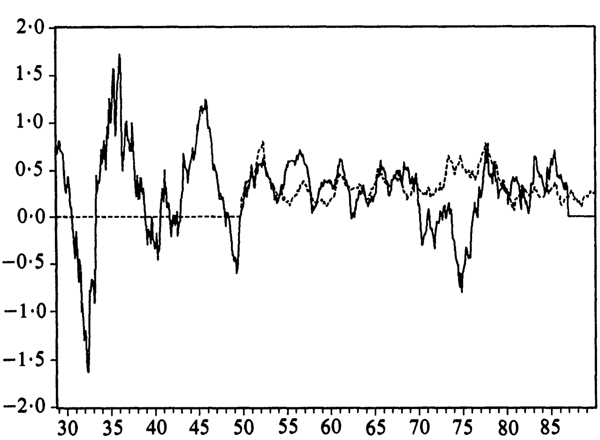

FIGURE 3(b)
3-year ahead predictions of 3-year EW returns: based on the recursive OLS updating procedure

portfolios. The plots suggest little success, and indeed the correlations are -0.4 and -0.05 respectively.

One way to interpret the failure of post-war prediction is that it confirms that evidence of mean reversion comes from the pre-war period. One sees the waning evidence of mean reversion in the sequence of monthly-updated estimates of the β for $K=3$ plotted in Figure 4 through the post-war period. The value-weighted β starts out about -0.65, loses a third of its value when the predicted market decline fails to occur in the 1950's, loses another third in the mid-1970's when the predicted rise fails to occur, and ends up at about -0.3. The equal-weighted β makes a fairly discrete jump toward zero in the mid-1970's as well.

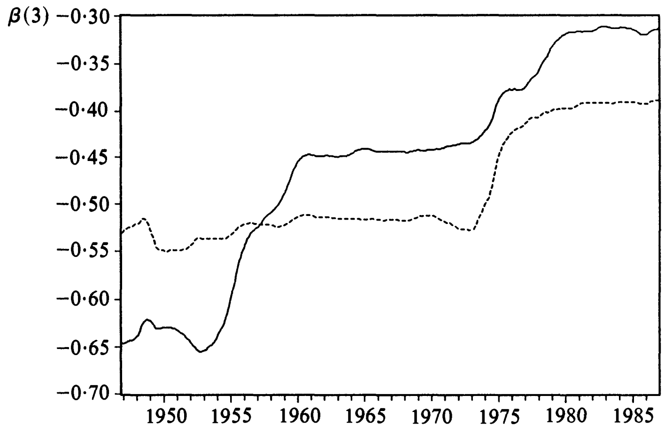

FIGURE 4

Successively updated $\beta(3)$: VW is solid line, EW is dotted line

4. CONCLUSIONS

This paper reappraises the evidence of mean reversion in stock market prices provided by the variance-ratio and multiperiod-return autoregression tests presented in a number of recent papers. Specifically, the variance ratio, which is theoretically one at any lag under the null hypothesis of a random walk, declines to below one at long lags in historical time series, and this has been interpreted as an indication of long-term mean-reverting behaviour in the stock market. Further it has been argued that 25–45% of the variation of 3–5 year holding-period stock returns is predictable from past returns, and that this is to be explained by the existence of a slowly decaying stationary component in stock prices. This paper challenges that view by re-examinating the evidence from different sub-periods and by using measures of statistical significance which do not depend on the assumption of normality.

By studying the impact of sample period on the test statistics we have shown mean reversion to be primarily a phenomenon of the 1926–1946 period which includes the Great Depression and World War II when the stock market was highly volatile. Mean reversion has not been a feature of the post-war era. On the contrary, post-war data displays, if anything, a tendency towards persistence in returns, or mean aversion, reflected in variance ratios that rise substantially above one at long lags and positive autoregression coefficients. Evidence of mean aversion after World War II is shown to be as strong as that for mean reversion over the whole period. Further, we find that the contrast between pre- and post-war results is unlikely under the hypothesis that returns were random over the whole period. We interpret these findings as evidence that the behaviour of stock returns changed at the end of World War II, perhaps because of the resolution of major uncertainties about the survival of the U.S. economy. To someone with the foresight to see that it would survive the Great Depression and fascism, the stock market was a bargain in the thirties and forties. Ex ante it may only have been an even bet. Ex post we see what looks like mean reverting behaviour, but this observation had no predictive power for the post-war environment. The transition to the post-war era of course did not occur at January 1947 or any specific date but must have occurred over a period of time. Predictions of a renewed depression were widespread at the end of the War, and these expectations must have influenced stock prices. As it became apparent that these predictions were wrong agents revised their estimates of parameters of the economy. Modelling this transition in the framework of learning under rational expectations as suggested by Bray and Savin (1986) might establish whether that process could account for the degree of mean aversion that we see in the post-war data.

The randomization method has been used here to develop standard errors and significance levels for test statistics which are free of distributional assumptions. There are several advantages to this computer-intensive method over Monte Carlo methods. Most important, it does not require that we pretend to know the underlying unknown distribution of stock market returns, but rather it focuses on testing the null hypothesis of randomness. Randomization is easy to execute and allows estimation of small sample distributions of test statistics which are often difficult to derive analytically. To isolate the effect of higher stock market volatility in the 1930s on the tests, we introduce a stratified randomization method. These experiments suggest that the significance of historical variance ratios has been overstated by Monte Carlo methods and by standard errors for overlapping autoregressions obtained by the method of Hansen and Hodrick.

Acknowledgement. We are grateful to W. C. Brainard, M. M. Bray, J. Y. Campbell, A. Harvey, S. H. Lee, A. W. Lo, J. M. Poterba, G. W. Schwert, L. H. Summers, and S. Wheatley for helpful comments, but responsibility for any errors is entirely the authors'. Nelson wishes to acknowledge support from the Center for Research in Banking and Security Markets.

REFERENCES

BARTLETT, M. S. (1946), "On the Theoretical Specification and Sampling Properties of Auto-correlated Time-Series", *Journal of the Royal Statistical Society*, **38**, 27-85.

BRAY, M. M. and SAVIN, N. E. (1986), "Rational Expectations Equilibria, Learning and Model Specification", *Econometrica*, **54**, 1129-1160.

CLARK, P. K. (1987), "Stationary Variation and Trend Reversion in U.S. Equity Prices" (mimeographed, Stanford University, Graduate School of Business).

COCHRANE, J. H. (1988), "How Big Is the Random Walk in GNP?", *Journal of Political Economy*, **96**, 893-920.

EFRON, B. (1979), "Bootstrap Methods: Another Look at the Jackknife", *The Annals of Statistics*, **7**, 1-26.

FAMA, E. F. (1970), "Efficient Capital Markets: A Review of Theory and Empirical Work", *Journal of Finance*, **25**, 383-416.

FAMA, E. F. (1990), "Term-Structure Forecasts of Interest Rates, Inflation, and Real Returns", *Journal of Monetary Economics*, **25**, 59-76.

FAMA, E. F. and FRENCH, K. R. (1988), "Permanent and Temporary Components of Stock Prices", *Journal of Political Economy*, **96**, 246-273.

FAMA, E. F. and FRENCH, K. R. (1989), "Business Conditions and Expected Returns on Stocks and Bonds", *Journal of Financial Economics*, **25**, 23-49.

FAUST, J. (1989), "A Variance Ratio Test for Mean Reversion: Statistical Properties and Implementation" (mimeographed, Federal Reserve Bank of Kansas City).

FISHER, R. A. (1935) *The Design of Experiments* (New York: Hafner Publishing Co).

HANSEN, L. P. and HODRICK, R. (1980), "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis", *Journal of Political Economy*, **88**, 829-853.

Stocks, Bonds, Bills, and Inflation: 1987 Yearbook (Chicago: Ibbotson Associates).

KENDALL, M. G. and STUART, A. (1976) *The Advanced Theory of Statistics* 3rd Edition, Vol. 3 (London: Griffin).

LEROY, S. F. (1982), "Expectations Models of Asset Prices: A Survey of Theory", *Journal of Finance*, **37**, 185-217.

LO, A. W. and 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.

LO, A. W. and 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.

NOREEN, E. (1989) *Computer-Intensive Methods for Testing Hypotheses: An Introduction* (New York: John Wiley & Sons).

OFFICER, R. R. (1973), "The Variability of the Market Factor of the New York Stock Exchange", *Journal of Business*, **46**, 434-453.

POTERBA, J. M. and SUMMERS, L. H. (1988), "Mean Reversion in Stock Returns: Evidence and Implications", *Journal of Financial Economics*, **22**, 27-59.

RICHARDSON, M. and STOCK, J. H. (1990), "Drawing Inferences from Statistics Based on Multi-Year Asset Returns" (NBER Working Paper No. 3335).

SCHWERT, G. W. (1989), "Why Does Stock Market Volatility Change Over Time?", *Journal of Finance*, **44**, 1115-1153.