NBER WORKING PAPER SERIES

MEAN REVERSION IN STOCK PRICES? A REAPPRAISAL OF THE EMPIRICAL EVIDENCE

Myung Jig Kim

Charles R. Nelson

Richard Startz

Working Paper No. 2795

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 1988

NBER Working Paper #2795
December 1988

MEAN REVERSION IN STOCK PRICES? A REAPPRAISAL OF THE EMPIRICAL EVIDENCE

ABSTRACT

Recent research based on variance ratios and multiperiod-return autocorrelations concludes that the stock market exhibits mean reversion in the sense that a return in excess of the average tends to be followed by partially offsetting returns in the opposite direction. Dividing history into pre-1926, 1926-46, and post-1946 subperiods, we find that the mean-reversion phenomenon is a feature of the 1926-46 period, but not of the post-1946 period which instead exhibits persistence of returns. Evidence for pre-1926 data is mixed.

The statistical significance of test statistics is assessed by estimating their distribution using stratified randomization. Autocorrelations of multiperiod returns imply a forecast of future returns, which is presented for post-war three-year returns using 1926-46, full sample, and sequentially updated coefficient estimates. The correlation between actual and forecasted returns is negative in each case.

We conclude that evidence of mean reversion in U.S. stock returns is substantially weaker than reported in the recent literature. If mean-reversion continues to be a feature of the stock market, then the experience of the past forty years has been an aberration.

Myung Jig Kim
Department of Economics DK-30
University of Washington
Seattle, WA 98195

Charles R. Nelson
Department of Economics DK-30
University of Washington
Seattle, WA 98195

Richard Startz
Department of Economics DK-30
University of Washington
Seattle, WA 98195

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 [1987] (hereafter P&S), Lo and MacKinlay [1987], 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, and should be unity under the random walk hypothesis. These authors find that historical variance ratios are below one at long lags, which they interpret as evidence of long term mean-reverting behavior in stock market prices. Cochrane showed that the variance ratio at lag K can be expressed as one plus a positively weighted sum of the sample autocorrelations of returns from lags one through K - 1. A pattern of variance ratios declining below unity as K increases therefore reflects negative long lag autocorrelation in returns. Evidently, a rise in prices is followed over time by a predictable decline. Using a closely related methodology based on autoregressions of multiperiod returns, Fama and French [1988] (hereafter F&F) claim to find evidence of mean-reversion and conclude that about 40% of the variation in stock returns is predictable from past returns.

This paper questions the finding of mean-reverting behavior in stock prices reached in the recent literature. Reexamination of the historical evidence for sample periods before and after the Great Depression and World War II suggests that the phenomenon identified in this literature is due primarily to that particular period. Post-war ratios do not in general display evidence of mean reversion. The autoregressions fitted by Fama and French have no predictive power after World War II. In addition, measures of statistical significance have been based on Monte Carlo simulations assuming Normal disturbances. Actual stock returns are generally recognized to be non-Normal. This paper presents estimates of the unknown distribution of the variance ratio and multiperiod autocorrelation statistics using the randomization method. The results suggest that historical variance ratios are not outside the range expected under the hypothesis that stock returns are random.

This paper is organized as follows. Section 2 briefly describes the background of this study and re-examines the historical variance ratios for annual data using the S&P Composite January average from Shiller [1981] (revised and updated) 1872-1986, and the December closing of the S&P Composite covering 1871-1987. We also study the monthly return series for value- and equal-weighted portfolios of NYSE stocks from the CRSP file for 1926-1986. Real and excess return versions of the data are considered. Section 3 discusses the randomization method for approximating the unknown sampling distribution of the test statistic and its exact significance level and present empirical results for the annual real S&P series. We also investigate the effect of the large variance of stock market returns in 1930s by introducing the idea of stratified randomization. Section 4 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 predictable is also attributable to the Depression-World War II era. Post-war prediction tests fail to reveal evidence of predictability. Section 5 concludes the paper.

2 Variance Ratios and Choice of Sample Period

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

$$VR(K) = \frac{var(r_t^K)}{var(r_t)} \cdot \frac{1}{K} \tag{1}$$

where r_t is the one period return, $r_t^K = \sum_{i=0}^{K-1} r_{t-i}$ and K is measured in years. The null hypothesis of a random walk is rejected if this statistic is significantly different from unity. Cochrane [1988] shows that VR(K) can be approximated by

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

where $\hat{\rho}(j)$ denotes the jth 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 jth sample autocorrelation has approximate expected value $-1/(T-j)$ as shown in Kendall and Stuart [1976], so that

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

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 the 12K-month return relative to the variation over one-year, as in Poterba and Summers [1987]. Following P&S we have

$$VR(k) = \frac{var(r_t^k)/k}{var(r_t^{12})/12} \tag{4}$$

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

where $r_t^k = \sum_{i=0}^{k-1} r_{t-i}$ and $\hat{\rho}(j)$ jth sample autocorrelation of monthly returns. P&S derive the expected value of VR(k) in (5) which provides a bias correction for monthly data.

For the annual historical data we use two standard data sets based on the Standard and Poor's Composite Index. One is the December closing value used by P&S which is available from 1871 from Wilson and Jones [1987] and Ibbotson [1987] with adjustment for inflation in real returns based on the CPI. The second is the January average value since 1872 using the PPI as the deflator as in Shiller [1981], updated by Shiller and ourselves through 1986. The estimated variance ratios over different sample periods for the two annual S&P series using formula (1) with bias correction (3) are reported in Tables 1 and 2 respectively. Our primary concern in this section is the effect of the choice of sample period on the results. The breaking points are 1926, which coincides with the starting date for the monthly data sets also used by P&S and by F&F, and 1947 which separates the period including the Great Depression and World War II from the post-war era.

The first three columns of Tables 1 and 2 are full sample variance ratios for nominal, real and excess returns. Our comments will focus on real and excess returns which are of greater economic interest than nominal returns. The basic mean-reversion result is evident in Figure 1 where the VR of real returns for each version of the S&P is plotted against K in years (for 1872-1986). Values below one show up immediately at low lags in the December closing data but not until lags in excess of ten years in the Shiller data. After that point both decline rapidly to about 0.4. Excess returns for the December closing show less mean reversion than real returns through lag eight, as reported by P&S. Beyond the eleven year lag neither excess return series displays mean-reversion. Monte Carlo standard errors from P&S are reported through lag eight years in Tables 1 and 2. We note that none of the sample VR's are more than about one standard error below one in this range of lags.

For the subsample period prior to 1926, the December closing returns display mean reversion for both real and excess returns. For the Shiller data this is the case only for excess returns, since VR's for real returns remain around one. During the subsample period 1926-1946 all versions of the VR decline sharply, though more so for the December closing data.

The post-war period starting in 1947 produces very different results, however. None of the six returns series in Tables 1 and 2 show evidence of mean-reversion at lags over five years, and then most notably for nominal returns. VR's for longer lags rise to about two for nominal returns and to about three for real and excess returns. It is clear that VR's below one for the whole history result from mixing the very low values prior to 1947, particularly the 1926-46 period, with high values after World War II. If mean-reversion continues to be a feature of the stock market, then the experience of the last forty years has been an aberration.

The next set of data we consider is CRSP monthly NYSE total returns, both equal- and value-weighted, which is available from 1926. The one-month T-bill rate and the CPI for all urban consumers (not seasonally adjusted) from Ibbotson Associates are used to calculate excess and real returns respectively. Table 3 shows the variance ratios calculated using the P&S formula (4) for sample periods 1926-86, 1926-46, and 1947-86. For the whole period, all of the return series show declining variance ratios. Using the Monte Carlo standard errors reported by P&S for series of length 720, one would conclude that VR is marginally significantly below unity at longer lags, particularly for the equal-weighted portfolio. The sample period 1926-46 which includes the Depression displays an even more severe decline in VR with increasing lag, as it did for the annual data. However, for the post-war period the story is very mixed. For the equal-weighted portfolio VR declines to about .7 at lag four years, but then rises and is above one at lag ten for real returns. In the case of the value-weighted portfolio, the VR drops below .7 for nominal returns at low lags, but nominal, real and excess return VR's all rise with lag until at lag ten years are well above one as in the annual data. Corresponding results excluding dividends are not reported here but are very similar.

One way to summarize the CRSP results is that the patterns are reversed when we compare the 1926-1946 period with the post-war period. During the earlier period VR's start above one at low lags and decline with increasing lag, while during the later period VR's are below one at low lags and then rise to well above one at long lags. In general, what evidence there is for mean reversion is stronger for equal-weighted returns than for value-weighted, as reported in previous studies.

3 Estimating Significance Levels for Variance Ratios

Since the small sample distribution of VR's has not been derived analytically, the existing literature reports standard errors estimated by Monte Carlo methods under the assumption of normality, or using Bartlett's formula through Cochrane's approximation to the VR. Cochrane [1988] presents standard errors from a Monte Carlo experiment using 100 observations of random walk with drift in his study of annual per capita real GNP series for the period 1869 - 1986. Drawing innovation errors from a standard normal distribution, he concludes that Bartlett's standard error only slightly understates the Monte Carlo standard errors at long lag and thus can be used as an approximation. P&S report Monte Carlo estimates of the standard deviation of VR under the normality assumption. In their test for mean reversion, F&F also draw innovation errors from a standard normal distribution. However, they also perform a Monte Carlo experiment where, in each replication, they simulate a heteroscedastic random walk that changes the standard deviation of the white-noise returns every 2 years to approximate variation through time in historical stock return variances. What all of these amount to is making implicit or explicit assumptions about the distribution of stock returns in order to assess the significance of test statistics. The nonnormality of stock returns is well-known. Figures 2, 3 and 4 plot of the frequency of annual and monthly real returns and the implied (by sample mean and standard deviation) normal distribution. Departure from normality seems to be most clear for the equal- and value-weighted CRSP real returns.

Both the bootstrap and randomization (or shuffling) methods are appropriate when the population distribution is unknown, since they both rely on resampling 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 $\mathbf{x} = (x_1, x_2, \cdots, x_T)$ from an unknown probability distribution $\mathbf{F}$, estimate the sampling distribution of some prespecified random variable $R(\mathbf{x}, \mathbf{F})$, on the basis of the observed data $\mathbf{x}$. Namely, we proceed as if the sample is the population for purposes of estimating the sampling distribution of the test statistic, $R(\mathbf{x}, \mathbf{F})$. 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 disturbances. The idea of randomization tests appeared in the literature as early as Fisher, R. A. [1935]. Noreen [1986] 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 the present context the null hypothesis 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.

The advantage of this approach over Monte Carlo is that the null hypothesis is very simple since no assumptions are made concerning the distribution of stock prices. Furthermore, as Noreen [1986] 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.[1]

Randomization estimates of the distribution of VR(K) for the real return based on the S&P Composite index 1871-1987 are reported in Table 4. Note that this is the series that seems to show most evidence

[1]For results reported below we randomize the order of the differenced series 1000 times. Specifically, each time we generate the random numbers bounded from 1 to T - 1 using GAUSS and pick up the corresponding row from the original return series. In bootstrapping same row may appear more than once, though with small probability.

of mean reversion for the whole historical period, the Shiller data and excess returns showing little or none. The historical sample values appear in the first column. Parameters of the sampling distribution such as mean, median and standard deviation (SD) of VR(K) estimated by randomization are reported as well as Bartlett's approximation of SD, SD^B for comparison. Also reported are the lower fractiles of the distribution and estimated significance levels, the probability that the variance ratios from a random sample is less than the historical value. Estimated means confirm that the correction for bias works reasonably well. The fact that medians are below means reflects the positive skewness of the sampling distribution of VR(17) that is evident in the histogram shown in Figure 5 against a reference normal distribution. This suggests that critical values based on fractiles and corresponding significance levels may differ importantly from those based on standard errors under a t-distribution approximation for VR.

Bartlett's approximation SD^B for the standard deviation closely approximates the estimated SD at short horizons, but understates the dispersion of VR by an amount that increases with lag. At lag 15 years. SD and SD^B are 0.410 and 0.286, respectively, and the discrepancy widens to .646 and .177 at lag 29. Since evidence of mean reversion rests on long lag VRs, understatement of sampling error as lag increases would seem to be a critical shortcoming of Bartlett's formula.

The last column of Table 5 gives the estimated significance level of VR, the probability of observing a VR as small as that observed in the historical data under the null hypothesis of no autocorrelation. It is only for lags of twenty three years or longer that significance levels fall below 10%. Note that these significance levels based on estimated fractiles are smaller than would be the case if we assumed a normal distribution for VR with the estimated standard deviation, and much larger than ones based on Bartlett's formula. P&S argue that " ⋯ unless one is strongly attached to the random walk hypothesis. significance levels in excess of .05 seem appropriate in evaluating the importance of transitory components in stock prices," and furthermore suggest that for the variance ratio test, a "40 % " significance level is appropriate if the goal is to minimize the sum of Type I and Type II errors. This criterion would reject the random returns hypothesis for any lag. Under more conventional significance criteria we would reject only if we had chosen to base the test on the longest lags. We note that results across lags are not independent and we do not have a portmanteau statistic one which to base a joint test. Bootstrapping yields the same substantive results.

Table 5 presents randomization results for the 1926-1986 subperiod using the annual December closing real total return S&P data. This period corresponds to the CRSP sample period. Estimated standard errors are, of course, larger for this shorter sample period. The estimated significance level of the historical VR is below P&S's 40% at lags below ten years and over 25 years. The smallest value is 19% and most are around 50%. Unless one is strongly inclined to find a transitory component in stock prices, these results offer little encouragement to that hypothesis relative to the random returns hypothesis. More generally, they reiterate the point that finding declining variance ratios in stock returns does not necessarily imply strong evidence against the random walk hypothesis. Bootstrapping again produces very similar results. We do not do the test for the post-war period which the reader will recall showed no evidence of mean reversion, but instead rising VR's.

As the investigation of historical variance ratios in section 2 suggests, fast declining variance ratios for the 1926 - 1946 period seems to account for much of the overall slowly decaying variance ratios for the whole historical sample period. A number of studies such as Officer [1973] and more recently Schwert [1988] have documented the much higher variance of stock prices during the 1930s. Studying market-factor variability over the period of 1897 to 1969, Schwert finds the decline of the variability of returns after 1930s is a return to levels prevailing before the Great Depression. Since the inclusion or exclusion of the Depression years gives rise to differing results it is natural to think of controlling for sampling period by using stratified randomization. Since one may reasonably suspect that stock returns in 1930s were drawn from an unknown distribution with higher variance, the distribution of the test statistic could be generated under the null hypothesis that within a given state of the world (the high variance state or the relatively low variance state) returns are stochastically independent. This is the same as putting historical returns into two different urns, depending upon whether they come from the 1930s or not, and shuffle independently to preserve the high variance property during 1930s in the randomization test.

Table 6 reports stratified randomization estimates of the sampling distribution of the VR for real S&P annual returns based on December closing for the 1871 to 1987 period. We put each historical return into one of two different urns depending on whether it belongs to the 1930-1939 period or not, and then shuffle independently, replicating 1000 times. The differences between the stratified sampling distributions and the corresponding ones in Table 4 are not negligible. For example, at lag 19 years the mean and median are both reduced by stratification and the exact significance of the historical sample value rises from 17% to 23% (i.e. is less significant). Stratification lowers the expected value of VR(19) by approximately 0.1 which implies that the low VRs for the Depression subsample cannot be attributed simply to high variance. Skewness in the sampling distribution of VR is apparent in the histogram of VR(19) in Figure 6. We have included longer lags in this experiment to see if the decline in significance levels with lag seen in Table 4 continues, and it does not.

What can we say at this point about the weight of evidence against the random return hypothesis and in favor of mean reversion in the stock market based on the annual data? We have estimated the exact significance levels for the set of historical VR's that are most favorable to the mean reversion alternative, namely real returns for the December closing S&P. Recall that VR's for the Shiller real returns do not drop below one until lags exceeding 15 years. VR's for excess returns from both data sets rise with lag to well above one (see Tables 1 and 2). We cannot assess the joint significance of VR's across a range of lags but only at a particular lag. If one came to the data with the expectation of finding evidence against randomness at lags in excess of twenty years, there is some if we include data prior to 1926 but not for the period 1926-87.

4 Evidence of Mean-Reversion Based on Autoregression of Multiperiod Returns

A close relative of variance ratios is the regression coefficient calculated by Fama and French [1987, 1988] and Fama [1988]. They regress the cumulative return from period t+1 to period t+K on the return from t-K+1 to t, so their estimating equation is

$$r_{K,t+K} = \alpha_K + \beta_K \; r_{K,t} + \epsilon_{K,t+K} \tag{6}$$

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

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

so both are functions of sample autocorrelations of the one-period 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.

F&F show that β is an estimate of the fraction of the variance of the K period return that is attributable to, and predictable from, a transitory component stock prices. Following Hansen and Hodrick [1980], they provide standard errors for the OLS slope which reflects the autocorrelation in residuals induced by overlapping observations. They also adjust for bias estimated by Monte Carlo simulation. F&F report strong evidence of mean-reversion at return horizons of three to five years in the monthly CRSP data, where 25% to 40% of the return is predictable from past returns.

Table 7 replicates some of the F&F results for the CRSP data 1926-86 for real and excess returns on the equal-weighted and value-weighted NYSE portfolios. Estimates of bias (based on Monte Carlo) and standard errors (based on Hansen and Hodrick [1980](hereafter H&H) for β from F&F appear in the lower panel. Even allowing for downward bias, β appears to be highly significant at lags of three to five years for equal- and value-weighted portfolios in real and excess return forms. The β estimate is also calculated for sub-samples 1926-46 (the Great Depression and World War II) and 1947-86. As in the VR results it is the 1926-46 period which accounts for the evidence of mean-reversion. While β is a large negative value in every case for the 1926-46 data, it is positive in most cases for the post-war data and particularly so at the three to five year lags on which F&F base their conclusions.

The randomization technique is used here to investigate the sampling distribution of the β statistic for the equal-weighted and value-weighted CRSP portfolios 1926-86. The null hypothesis to be tested is 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 the coefficient we repeat the procedure many times.

Randomization results for real returns on the equal-weighted NYSE portfolio CRSP for 1926-86 are presented in Table 8. 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 considerably larger than the H&H SD estimates at lags three to six years reported in Table 7. Estimated significance levels are as small as 2.8% (at lag 4 years), but exceed 30% at all but lags two through five years. Again, we lack a measure of joint significance across lags. Unlike the case of VR's, we get similar significance levels if we use the estimated bias and SD to calculate a t-statistic, as shown in the first two lines of the bottom panel of Table 8. The last two lines show that much smaller significance levels are implied (incorrectly) by H&H SD.

Corresponding randomization results are reported in Table 9 for the real returns on the value-weighted NYSE portfolio. Recall that the evidence of mean reversion was weaker for value-weighted portfolios, with bias-corrected βs only about one standard deviation from zero. Significance levels based on randomization are only as low as about 10% at lags two and three years and range as high as about 60% at lags five and six. Again, we have no basis for combining significance levels across lags. It seems clear however that to find evidence of mean reversion in the value-weighted returns would require a prior choice of particular lags for the test. Finally, we note here that the obvious advantage of the randomization method in this case is simplicity. All we need to do is to run OLS, not worrying about the method of adjustment of standard deviations such as the one of H&H which evidently produces SD's which are much too small, leading to incorrect inferences of significance.

Stratified randomization can again be used to allow for a change in the distribution of returns during the 1930's. The results for equal-weighted NYSE portfolio returns, those most favorable to mean reversion, are shown in Table 10. Comparing significance levels with those in Table 8, Stratification has the effect of raising them at all lags. At lags two through five where p-values were well below 10% in Table 8, they are roughly tripled. Stratification of the sample therefore substantially weakens evidence for mean-reversion, even in the case of equal-weighted returns.

The F&F approach lends itself directly to post-sample testing since it provides a predictive relation for future returns. Given an estimate of β (and the intercept), the previous K-year return implies the predicted value of the next. In the first experiment we use the OLS values estimated for 1926-46 to forecast from each month the return on the value-weighted portfolio for the next three years, successively through the post-war period. The resulting forecasts and actuals are plotted in the top panel of Figure 7. One get little impression of predictive ability for past returns, indeed the correlation between actuals and predictions is - 0.08. In the lower panel we use the sample estimates based on the whole period 1926-86, those which gave the best fit after the fact. The correlation is again - 0.08. To simulate real-time forecasting we also forecasted ahead each month based on coefficients estimated up through that point in time and the results are plotted in Figure 8 for both equal-weighted and value-weighted portfolios. The plots suggest little relation between actuals and predictions, indeed the correlations are -.4 and - 0.05 respectively.

One way to interpret the failure of post-war prediction is that it confirms the fact that evidence of mean-reversion comes from the pre-war period. One sees the waning evidence of mean-reversion in the sequential estimates of the βs for K = 3 plotted in Figure 9 through the post-war period. The value-weighted β starts out about .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 .3. The equal-weighted β makes a fairly discrete jump toward zero in the mid-1970's as well.

5 Conclusion

This paper reappraises the evidence of mean reversion in stock market prices provided by the variance ratio and related 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. This has been interpreted as an indication of long-term mean-reverting behavior in the stock market. Further it has been argued that 25 to 40 percent of the variation of 3 to 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. The purpose of this paper is to challenge this view, based on re-examination of the evidence from different sub-periods and measures of 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 1926-46 period which includes the Great Depression and World War II when the stock market was highly volatile, plunging several times and then recovering. 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 reflected in variance ratios that rise substantially above one at long lags and in positive multiyear autoregression coefficients.

The randomization method is used to develop 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 1930's on the tests, we introduce a stratified randomization method which essentially puts those observations in a separate urn.

Stratified randomization of annual real total returns on the S&P Composite Index since 1871 suggests that significance levels are in the range of 20 to 30% for individual lags up to 20 years, and around 10% for lags of 25 years and longer. For monthly returns 1926-1986 we randomize to estimate the distribution of the multiperiod return autoregression coefficient. We obtain significance level as low as 3% for equal-weighted portfolios of NYSE stocks and 9% for value-weighted portfolios at a lag of three years, but significance levels are as high as 70% at longer lags. Further, stratification of the sample raises significance levels to 9% at best even for equal-weighted returns. In general we find that standard errors reported in previous studies based on other methods would imply much stronger significance.

To sum up, while evidence of mean-reversion arises only from pre-World War II data, the magnitude of test statistics for larger sample periods is not readily dismissed purely as the result of chance. On the other hand, to build a strong case for mean-reversion one would have to argue the choice of particular data sets and particular lags on prior grounds: real returns rather than excess returns, lags in excess of twenty years for annual data but around four years for monthly data, and equal-weithted portfolios rather than value-weighted. One would also have to settle for significance levels well above the customary 5%. The need for measures of joint significance across lags is clear. If one rejects the random returns hypothesis, it must be on the basis of the behavior of stock returns during the 1926-46 period. The alternative hypothesis that has been suggested is predictability of returns resulting from a transitory component of stock prices. Another alternative is suggested by the historical context of the 1930's and 40's. This was a period during which the possibility that the U.S. economy would disintegrate recurrently boomed large and then receeded. A priori it was not obvious that the U.S. corporate system would survive the Depression or World War II, but a posteriori it did and perhaps has left us with an apprarent episode of mean-reversion.

References

[1] Clark, P. K. "Stationary Variation and Trend Reversion in U.S. Equity Prices ". Mimeographed. Stanford University, Graduate School of Business, July 1987.

[2] Cochrane, J. H. "How Big Is the Random Walk in GNP? ". *J. Political Economy.* 96. (October 1988). 893-920

[3] Efron, B. "Bootstrap Methods: Another Look at the Jackknife ". *The Annals of Stat.*. 7. (1979) 1-26.

[4] Fama, E. F. "Term Structure Forecasts of Interest Rates, Inflation. and Real Returns ". Mimeographed. UCLA, January 1988.

[5] __________. "Efficient Capital Markets: A Review of Theory and Empirical Work ". *J. Finance.* 25 (May 1970), 383-416.

[6] Fama, E. F. and K. R. French. "Forecasting Returns on Corporate Bonds and Common Stocks ". Mimeographed. University of Chicago, December 1987.

[7] __________. "Permanent and Temporary Components of Stock Prices ". *J. Political Economy.* 96 (April 1988): 246-73.

[8] Fisher, R. A. *The Design of Experiments.* Hafner Publishing Co. New York, 1935.

[9] Gordon, R. J. *The American Business Cycle.* The Univ. of Chicago Press. Chicago, 1986.

[10] Hansen, L. P. and R. Hodrick. "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis ". *J. Political Economy.* 88, (1980), 829-53.

[11] Ibbotson Associates. *"Stocks, Bonds, Bills, and Inflation: 1987 Yearbook"*. Chicago: Ibbotson Associates, 1987.

[12] Kendall, M. G. and A. Stuart. *"The Advanced Theory of Statistics "*. 3rd Edition, Vol. 3. London: Griffin, 1976.

[13] LeRoy, S. F. "Expectations Models of Asset Prices: A Survey of Theory". *J. Finance*. 37, (March 1982), 185-217.

[14] Lo, A. W. and A. C. MacKinlay. "A Simple Specification Test of the Random Walk Hypothesis". Rodney L. White Center for Financial Research Discussion Paper. 13-87, Wharton School, University of Pennsylvania, 1987.

[15] Noreen, E. *An Introduction to Testing Hypotheses Using Computer-Intensive Methods*. Unpublished. Univ. of Washington, Seattle, 1986.

[16] Officer, R. R. "The Variability of the Market Factor of the New York Stock Exchange". *J. Business*. 46, (1973), 434-53.

[17] Poterba, J. M. and L. H. Summers. "Mean Reversion in Stock Returns: Evidence and Implications". Mimeographed, Cambridge, Mass.: NBER. March 1987.

[18] Shiller, R. J. "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?". *American Econ. Review*. 71 (June 1981), 421-36.

[19] Wilson, J. W. and C. Jones. "A Comparison of Annual Common Stock Returns: 1871-1925 with 1926-1985". *J. Business*. 60 (April 1987), 239-58.

</ref_section>

Table 1: Sample Variance Ratios for Annual S&P Composite Index Total Returns, December Closing

	1871 - 1987			1871 - 1925			1926 - 1946			1947 - 1987		
K	(N)	(R)	(E)	(N)	(R)	(E)	(N)	(R)	(E)	(N)	(R)	(E)
2	1.037	1.027	1.056	0.946	1.023	0.957	1.248	1.093	1.252	0.949	1.032	1.005
	(.095)			(.140)								
3	0.894	0.876	0.936	0.650	0.797	0.667	1.308	1.058	1.329	0.713	0.876	0.842
	(.143)			(.210)								
4	0.865	0.849	0.932	0.624	0.818	0.644	1.251	0.930	1.303	0.682	0.907	0.881
	(.179)			(.265)								
5	0.827	0.827	0.922	0.629	0.850	0.648	0.976	0.649	1.057	0.856	1.121	1.104
	(.211)			(.313)								
6	0,742	0.773	0.867	0.481	0.727	0.500	0.718	0.447	0.814	1.011	1.329	1.312
	(.240)			(.358)								
7	0.679	0.752	0.831	0.440	0.725	0.453	0.527	0.340	0.605	1.051	1.451	1.420
	(.266)			(.398)								
8	0.664	0.770	0.841	0.466	0.788	0.468	0.390	0.278	0.440	1.141	1.620	1.590
	(.290)			(.436)								
9	0.631	0.746	0.841	0.406	0.770	0.397	0.368	0.212	0.412	1.201	1.753	1.732
10	0.618	0.722	0.869	0.356	0.714	0.349	0.451	0.292	0.471	1.257	1.883	1.863
11	0.639	0.720	0.934	0.384	0.718	0.382	-	-	-	1.309	2.020	1.995
12	0.668	0.726	1.006	0.386	0.721	0.389	-	-	-	1.388	2.180	2.165
13	0.668	0.717	1.053	0.338	0.678	0.346	-	-	-	1.431	2.304	2.292
14	0.646	0.701	1.082	0.324	0.660	0.335	-	-	-	1.477	2.447	2.461
15	0.638	0.692	1.114	0.331	0.668	0.342	-	-	-	1.580	2.607	2.641
20	0.573	0.529	1.233	0.185	0.462	0.236	-	-	-	1.969	3.051	3.197
25	0.541	0.372	1.352	0.234	0.384	0.313	-	-	-	-	-	-
30	0.539	0.303	1.469	-	-	-	-	-	-	-	-	-

Note:
1. Parentheses are Monte Carlo standard errors, based on 115 and 55 observations, Reported in Porterba and Summers [1987]. (N), (R), and (E) denote Nominal, Real, and Excess returns, respectively.
2. Variance ratios for annual returns are calculated from

$$VR(K) = \frac{var(r_t^K)/K}{var(r_t)}$$

where r_t^K = K-holding period returns and $K = 1, 2, \cdots$(year). To correct the downward-bias sample variance ratio is divided by the expected value of VR(K) given in (3).
Data: see Data Description in the Appendix.

Table 2: Sample Variance Ratios for Annual S&P Composite Index Total Returns, Shiller's January Average

	1872 - 1986			1872 - 1925			1926 - 1946			1947 - 1986		
K	(N)	(R)	(E)	(N)	(R)	(E)	(N)	(R)	(E)	(N)	(R)	(E)
2	1.081	1.074	1.112	1.009	1.097	1.023	1.273	1.161	1.266	0.959	1.070	1.067
	(.095)			(.140)								
3	0.964	0.998	1.027	0.770	1.025	0.802	1.368	1.222	1.356	0.765	0.957	0.975
	(.143)			(.210)								
4	0.944	1.015	1.037	0.696	1.062	0.724	1.414	1.296	1.411	0.755	1.006	1.071
	(.179)			(.265)								
5	0.881	1.011	1.010	0.677	1.087	0.701	1.108	1.024	1.127	0.891	1.168	1.326
	(.211)			(.313)								
6	0.785	1.005	0.949	0.544	1.031	0.578	0.843	0.822	0.883	1.063	1.360	1.591
	(.240)			(.358)								
7	0.715	1.013	0.916	0.467	1.039	0.499	0.701	0.655	0.753	1.111	1.496	1.764
	(.266)			(.398)								
8	0.693	1.056	0.927	0.463	1.057	0.482	0.620	0.644	0.648	1.217	1.689	2.008
	(.290)			(.436)								
9	0.667	1.065	0.940	0.448	1.069	0.459	0.590	0.723	0.572	1.314	1.841	2.272
10	0.663	1.069	0.975	0.404	1.042	0.417	0.740	0.839	0.713	1.393	1.980	2.460
11	0.687	1.079	1.045	0.401	1.045	0.416	-	-	-	1.450	2.117	2.621
12	0.707	1.062	1.118	0.418	1.068	0.438	-	-	-	1.481	2.239	2.819
13	0.714	1.032	1.174	0.388	1.060	0.419	-	-	-	1.568	2.375	3.049
14	0.705	0.985	1.204	0.371	1.070	0.408	-	-	-	1.626	2.460	3.168
15	0.695	0.938	1.232	0.376	1.092	0.413	-	-	-	1.702	2.556	3.322
20	0.624	0.717	1.353	0.211	0.940	0.297	-	-	-	1.953	2.783	3.931
25	0.584	0.450	1.470	0.244	1.060	0.355	-	-	-	-	-	-
30	0.579	0.337	1.600	-	-	-	-	-	-	-	-	-

Note:

1. Parentheses are Monte Carlo standard errors, based on 115 and 55 observations, Reported in Porterba and Summers [1987]. **(N)**, **(R)**, and **(E)** denote Nominal, Real, and Excess returns, respectively.
2. Variance ratios for annual returns are calculated from

$$VR(K) = \frac{var(r_t^K)/K}{var(r_t)}$$

where r_t^K = K-holding period returns and $K = 1, 2, \cdots$(year). To correct the downward-bias sample variance ratio is divided by the expected value of VR(K) given in (3).

Data: see Data Description in the Appendix.

Table 2: Sample Variance Ratios for Annual S&P Composite Index Total Returns, Shiller's January Average

	1872 - 1986			1872 - 1925			1926 - 1946			1947 - 1986		
K	(N)	(R)	(E)	(N)	(R)	(E)	(N)	(R)	(E)	(N)	(R)	(E)
2	1.081	1.074	1.112	1.009	1.097	1.023	1.273	1.161	1.266	0.959	1.070	1.067
	(.095)			(.140)								
3	0.964	0.998	1.027	0.770	1.025	0.802	1.368	1.222	1.356	0.765	0.957	0.975
	(.143)			(.210)								
4	0.944	1.015	1.037	0.696	1.062	0.724	1.414	1.296	1.411	0.755	1.006	1.071
	(.179)			(.265)								
5	0.881	1.011	1.010	0.677	1.087	0.701	1.108	1.024	1.127	0.891	1.168	1.326
	(.211)			(.313)								
6	0.785	1.005	0.949	0.544	1.031	0.578	0.843	0.822	0.883	1.063	1.360	1.591
	(.240)			(.358)								
7	0.715	1.013	0.916	0.467	1.039	0.499	0.701	0.655	0.753	1.111	1.496	1.764
	(.266)			(.398)								
8	0.693	1.056	0.927	0.463	1.057	0.482	0.620	0.644	0.648	1.217	1.689	2.008
	(.290)			(.436)								
9	0.667	1.065	0.940	0.448	1.069	0.459	0.590	0.723	0.572	1.314	1.841	2.272
10	0.663	1.069	0.975	0.404	1.042	0.417	0.740	0.839	0.713	1.393	1.980	2.460
11	0.687	1.079	1.045	0.401	1.045	0.416	-	-	-	1.450	2.117	2.621
12	0.707	1.062	1.118	0.418	1.068	0.438	-	-	-	1.481	2.239	2.819
13	0.714	1.032	1.174	0.388	1.060	0.419	-	-	-	1.568	2.375	3.049
14	0.705	0.985	1.204	0.371	1.070	0.408	-	-	-	1.626	2.460	3.168
15	0.695	0.938	1.232	0.376	1.092	0.413	-	-	-	1.702	2.556	3.322
20	0.624	0.717	1.353	0.211	0.940	0.297	-	-	-	1.953	2.783	3.931
25	0.584	0.450	1.470	0.244	1.060	0.355	-	-	-	-	-	-
30	0.579	0.337	1.600	-	-	-	-	-	-	-	-	-

Note:

1. Parentheses are Monte Carlo standard errors, based on 115 and 55 observations, Reported in Porterba and Summers [1987]. **(N)**, **(R)**, and **(E)** denote Nominal, Real, and Excess returns, respectively.
2. Variance ratios for annual returns are calculated from

$$VR(K) = \frac{var(r_t^K)/K}{var(r_t)}$$

where r_t^K = K-holding period returns and $K = 1, 2, \cdots$(year). To correct the downward-bias sample variance ratio is divided by the expected value of VR(K) given in (3).

Data: see Data Description in the Appendix.

Table 3: Variance Ratios of Monthly Total Returns for All NYSE Stocks from CRSP

K (years)	2	3	4	5	6	7	8	9	10
					time period: 1926 - 1986				
Equal-weighted									
Nominal	1.003	.910	.859	.752	.603	.426	.346	.317	.315
Real	.963	.833	.754	.644	.524	.399	.352	.330	.332
Excess	1.009	.923	.877	.783	.646	.487	.427	.421	.445
Value-weighted									
Nominal	1.032	.969	.898	.808	.715	.599	.557	.559	.585
Real	.977	.870	.767	.678	.619	.572	.581	.603	.642
Excess	1.035	.980	.919	.849	.775	.682	.671	.709	.771
MC SD	(.108)	(.177)	(.232)	(.278)	(.320)	(.358)	(.394)	(-)	(-)
					time period: 1926 - 1946				
Equal-weighted									
Nominal	1.129	1.124	1.166	1.020	.758	.452	.319	.290	.382
Real	1.050	.986	.990	.839	.607	.352	.245	.199	.222
Excess	1.141	1.154	1.216	1.078	.816	.494	.351	.319	.414
Value-weighted									
Nominal	1.175	1.205	1.157	.965	.731	.511	.434	.431	.551
Real	1.059	1.010	.909	.711	.505	.344	.295	.276	.322
Excess	1.182	1.231	1.210	1.037	.810	.568	.474	.460	.574
					time period: 1947 - 1986				
Equal-weighted									
Nominal	.889	.731	.689	.782	.879	.855	.823	.822	.803
Real	.918	.772	.732	.820	.934	.950	.965	1.005	1.030
Excess	.891	.739	.691	.784	.882	.876	.875	.902	.917
Value-weighted									
Nominal	.864	.692	.664	.781	.947	1.006	1.048	1.123	1.188
Real	.956	.864	.897	1.059	1.283	1.428	1.565	1.720	1.861
Excess	.924	.826	.851	1.026	1.239	1.366	1.496	1.649	1.778

Note:

1. Variance ratios are estimated using sample variance ratio formula (4) for monthly returns. To correct the downward-bias the sample variance ratio is divided by the expected value of VR(k) given by Porterba and Summers [1987].
2. Monte Calro standard errors are reported in parentheses from P&S for T = 720, namely for 1926-1985.

Table 4: The Randomization Sampling Distribution of Variance Ratios for Annual Real Total Returns, December Closing S&P Composite Index, 1871 - 1987.

						Fractiles†					Signif.
K	$VR(K)$	mean	median	SD	SD^B	1%	2.5%	5%	10%	20%	Level‡
2	1.027	-	-	-	.155	-	-	-	-	-	-
3	0.876	1.005	1.004	.138	.162	.706	.742	.784	.828	.884	17.9
4	0.847	-	-	-	.181	-	-	-	-	-	-
5	0.827	1.005	0.987	.203	.197	.601	.647	.690	.745	.832	19.3
6	0.773	-	-	-	.202	-	-	-	-	-	-
7	0.752	1.002	0.973	.254	.212	.515	.567	.628	.696	.786	16.2
8	0.770	-	-	-	.233	-	-	-	-	-	-
9	0.746	0.997	0.963	.299	.239	.447	.510	.567	.639	.743	20.5
10	0.722	-	-	-	.244	-	-	-	-	-	-
11	0.720	0.995	0.949	.337	.255	.400	.468	.532	.610	.712	21.0
12	0.726	-	-	-	.268	-	-	-	-	-	-
13	0.717	0.994	0.936	.373	.276	.365	.422	.501	.577	.673	24.0
14	0.701	-	-	-	.280	-	-	-	-	-	-
15	0.692	0.992	0.919	.410	.286	.369	.416	.467	.535	.643	24.4
17	0.612	0.992	0.911	.447	.269	.341	.384	.429	.502	.627	18.3
19	0.573	0.993	0.898	.483	.267	.285	.361	.401	.481	.606	17.3
21	0.501	0.995	0.886	.517	.245	.279	.341	.389	.457	.578	14.0
23	0.431	0.997	0.873	.551	.221	.261	.313	.369	.433	.560	9.8
25	0.372	1.000	0.860	.584	.199	.234	.287	.350	.420	.534	6.3
27	0.320	1.003	0.849	.616	.178	.222	.265	.330	.400	.518	4.4
29	0.309	1.006	0.848	.646	.177	.211	.256	.315	.383	.500	4.6
30	0.303	-	-	-	.177	-	-	-	-	-	-

Note:

1. We apply sample variance ratio formula, corrected for the bias.
2. SD^B reports Bartlett's approximation of standard error for VR(K), i.e., $SD^B \simeq \sqrt{4K/3T} \cdot VR(K)$ as is shown in Cochrane [1988].
3. † : reports lower x % fractiles of VR(K) from the estimated randomization sampling distribution.
4. ‡ : reports the probabilities that variance ratios from randomized samples are less than historical value of variance ratios.

Table 5: The Randomization Sampling Distribution of Variance Ratios for Annual Real Total Returns, December Closing S&P Composite Index, 1926 - 1987.

						Fractiles†					Signif.
K	$VR(K)$	mean	median	SD	SD^B	1%	2.5%	5%	10%	20%	Level‡
2	1.050	-	-	-	.218	-	-	-	-	-	-
3	0.920	0.993	0.979	.192	.234	.607	.653	.689	.747	.830	37.1
4	0.817	-	-	-	.240	-	-	-	-	-	-
5	0.737	0.995	0.965	.290	.242	.489	.523	.570	.646	.744	19.0
6	0.708	-	-	-	.254	-	-	-	-	-	-
7	0.690	0.995	0.929	.374	.268	.355	.425	.481	.577	.681	21.1
8	0.721	-	-	-	.299	-	-	-	-	-	-
9	0.742	1.000	0.916	.446	.326	.308	.361	.434	.505	.622	32.8
10	0.794	-	-	-	.368	-	-	-	-	-	-
11	0.865	1.007	0.892	.511	.421	.280	.320	.380	.449	.592	47.7
12	0.929	-	-	-	.472	-	-	-	-	-	-
13	0.985	1.019	0.891	.570	.521	.241	.288	.332	.427	.558	57.2
14	1.000	-	-	-	.549	-	-	-	-	-	-
15	0.987	1.030	0.884	.625	.561	.203	.247	.315	.401	.518	57.2
17	0.963	1.045	0.873	.681	.582	.199	.236	.283	.374	.494	54.9
19	0.997	1.063	0.855	.736	.637	.194	.232	.269	.350	.485	56.8
21	0.940	1.085	0.868	.792	.632	.181	.224	.270	.355	.467	54.0
23	0.883	1.108	0.879	.844	.621	.181	.215	.260	.338	.449	50.3
25	0.764	1.138	0.892	.889	.560	.162	.207	.249	.324	.453	43.0
27	0.661	1.175	0.909	.929	.504	.153	.196	.246	.322	.451	35.0
29	0.614	1.223	0.909	.971	.485	.171	.195	.239	.319	.459	31.4
30	0.576	-	-	-	.463	-	-	-	-	-	-

Note:

1. We apply sample variance ratio formula, corrected for the bias.
2. SD^B reports Bartlett's approximation of standard error for VR(K), i.e., $SD^B \simeq \sqrt{4K/3T} \cdot VR(K)$ as is shown in Cochrane [1988].
3. † : reports lower x % fractiles of VR(K) from the estimated randomization sampling distribution.
4. ‡ : reports the probabilities that variance ratios from randomized samples are less than historical value of variance ratios.

Table 6: Stratified Randomization Sampling Distribution of Variance Ratios for Annual Real Total Returns, December Closing S&P Composite Index, 1871 - 1987.

K	$VR(K)$	mean	median	SD	SD^B	Fractiles† 1%	2.5%	5%	10%	20%	Signif. Level‡
4	0.849	0.959	0.943	.178	.181	.618	.660	.698	.742	.805	29.9
7	0.752	0.927	0.894	.239	.212	.484	.553	.594	.655	.733	23.4
10	0.722	0.903	0.859	.284	.244	.405	.464	.520	.592	.669	28.5
13	0.717	0.891	0.836	.327	.276	.326	.393	.467	.530	.613	33.6
16	0.654	0.887	0.824	.369	.279	.300	.371	.418	.493	.571	29.6
19	0.573	0.884	0.808	.412	.267	.259	.318	.373	.441	.546	23.5
22	0.474	0.884	0.793	.453	.237	.250	.292	.336	.398	.514	16.0
25	0.372	0.887	0.772	.492	.199	.235	.273	.311	.382	.483	9.3
28	0.315	0.892	0.765	.531	.178	.221	.244	.286	.357	.459	6.5
31	0.319	0.898	0.759	.571	.190	.179	.243	.278	.343	.444	7.7
34	0.327	0.904	0.734	.610	.203	.179	.221	.262	.332	.423	9.6
37	0.337	0.913	0.731	.647	.219	.181	.214	.252	.316	.410	12.2
40	0.342	0.926	0.732	.684	.231	.176	.210	.246	.304	.401	13.8
43	0.317	0.941	0.736	.717	.222	.165	.202	.247	.297	.391	12.1
46	0.318	0.963	0.753	.748	.230	.158	.197	.242	.292	.383	12.5
49	0.289	0.988	0.759	.776	.216	.160	.203	.233	.283	.376	10.4

Note:

1. We use bias-corrected sample variance ratio formula in this experiment.
2. The 1930-1939 period is placed in different urn.
3. SD^B: reports Bartlett's approximation of standard error for VR(K), i.e., $SD^B \simeq \sqrt{4K/3T} \cdot VR(K)$.
4. †: reports lower x % fractiles of VR(K) from estimated randomization sampling distribution.
5. ‡: reports the probabilities that variance ratios from randomized samples are less than historical value of variance ratios.

Table 7: OLS Slopes for the Monthly Equal- and Value-Weighted NYSE Total Returns from CRSP (both CPI-Adjusted and Excess Returns)

$$r_{K,t+K} = \alpha_K + \beta_K \, r_{K,t} + \epsilon_{K,t+K}$$

		Return Horizon (K: years)									
		1	2	3	4	5	6	7	8	9	10
		OLS Slopes (β_K): Equal-Weighted									
Real											
	1926-1946	-.04	-.23	-.53	-.67	-.72	-.39	.22	-.83	-.41	-1.89
	1947-1986	-.12	-.27	.06	.09	-.05	-.21	-.25	-.28	-.34	-.43
	1926-1986	-.07	-.26	-.39	-.47	-.47	-.28	-.10	-.16	-.11	-.10
Excess											
	1926-1946	.04	-.13	-.45	-.59	-.56	-.17	.29	-.07	-.44	-1.44
	1947-1986	-.15	-.29	.05	.05	-.10	-.22	-.21	-.21	-.25	-.31
	1926-1986	-.02	-.18	-.32	-.40	-.37	-.15	.09	.10	.11	.11
		OLS Slopes (β_K): Value-Weighted									
Real											
	1926-1946	-.02	-.26	-.65	-.64	-.61	-.28	.23	-1.12	-.29	-1.97
	1947-1986	-.07	-.13	.33	.48	.38	.23	.20	.20	.12	.05
	1926-1986	-.05	-.23	-.31	-.20	-.07	-.09	.17	.07	.03	-.08
Excess											
	1926-1946	.08	-.15	-.54	-.54	-.40	-.05	.27	-1.06	-.52	-1.76
	1947-1986	-.11	-.16	.34	.50	.41	.28	.25	.24	.17	.10
	1926-1986	.01	-.15	-.24	-.17	-.03	.16	.30	.27	.20	.10
		Mean-Bias Adjustment Factors of OLS Slopes and SDs 1926 - 1985 Source: Fama and French [1988]									
MC mean		-.02	-.04	-.07	-.10	-.13	-.16	-	-.22	-	-.29
H&H SD(Equal)†		.11	.14	.14	.13	.14	.17	-	.24	-	.28
H&H SD(Value)‡		.11	.14	.15	.16	.20	.23	-	.34	-	.43
SD(MC Hetero.)		.15	.19	.20	.22	.23	.25	-	.25	-	.28

†: Standard errors reported in Fama and French [1988] for *equal-weighted* NYSE market portfolios for period 1926 - 1985. The standard errors of the OLS slopes are adjusted for the residual autocorrelation due to overlap of monthly observations on longer-horizon returns with the method of Hansen and Hodrick [1980].

‡: Standard errors reported in Fama and French [1988] for *value-weighted* NYSE market portfolios for period 1926 - 1985, using the method of H&H.

Table 8: Randomization Estimates of OLS Slope Sampling Distribution for Monthly Equal-Weighted NYSE Total Returns (CPI-Adjusted) from CRSP: 1926 - 1986

$$r_{K,t+K} = \alpha_K + \beta_K \; r_{K,t} + \epsilon_{K,t+K}$$

	Return Horizon (K: years)									
	1	2	3	4	5	6	7	8	9	10
	Historical OLS Slopes (β_K)									
	-.07	-.26	-.39	-.47	-.47	-.28	-.10	-.16	-.11	-.10
	Randomization Results									
mean	-.02	-.04	-.06	-.09	-.12	-.15	-.17	-.20	-.23	-.27
median	-.02	-.03	-.08	-.10	-.13	-.15	-.17	-.21	-.25	-.28
SD	.10	.15	.18	.21	.23	.24	.26	.28	.30	.31
	Fractiles									
1%	-.24	-.40	-.45	-.54	-.58	-.69	-.72	-.80	-.89	-.96
2.5%	-.21	-.34	-.42	-.49	-.53	-.59	-.66	-.75	-.79	-.86
5%	-.19	-.29	-.37	-.42	-.48	-.53	-.60	-.66	-.72	-.75
10%	-.16	-.23	-.30	-.36	-.41	-.47	-.52	-.57	-.61	-.65
20%	-.11	-.17	-.22	-.27	-.32	-.35	-.41	-.45	-.49	-.52
	Significance Level									
	30.7%	6.9%	3.9%	2.8%	5.9%	30.0%	60.3%	56.1%	66.7%	70.7%
	t-values and One-sided Significance Level									
t-value	-.50	-1.47	-1.83	-1.81	-1.52	-.54	-	-	-	-
from RAND	30.9%	7.1%	3.4%	3.5%	6.4%	29.5%	-	-	-	-
t-value	-.45	-1.57	-2.36	-2.85	-2.43	-.71	-	-	-	-
implied by H&H	32.6%	5.8%	0.9%	0.2%	0.7%	23.9%	-	-	-	-

Table 9: Randomization Estimates of OLS Slope Sampling Distribution for Monthly Value-Weighted NYSE Total Returns (CPI-Adjusted) from CRSP: 1926 - 1986

$$r_{K,t-K} = \alpha_K + \beta_K \, r_{K,t} + \epsilon_{K,t+K}$$

	Return Horizon (K: years)									
	1	2	3	4	5	6	7	8	9	10
	Historical OLS Slopes (β_K)									
	-.05	-.23	-.31	-.20	-.07	-.09	.17	.07	.03	-.08
	Randomization Results									
mean	-.02	-.04	-.07	-.10	-.12	-.15	-.18	-.21	-.24	-.26
median	-.02	-.04	-.07	-.10	-.12	-.15	-.18	-.22	-.24	-.28
SD	.10	.15	.18	.20	.22	.24	.26	.28	.29	.31
	Fractiles									
1%	-.26	-.39	-.46	-.54	-.62	-.67	-	-	-	-
2.5%	-.21	-.36	-.42	-.48	-.53	-.61	-	-	-	-
5%	-.19	-.29	-.36	-.42	-.48	-.55	-	-	-	-
10%	-.16	-.24	-.31	-.36	-.42	-.47	-	-	-	-
20%	-.11	-.17	-.22	-.27	-.32	-.36	-	-	-	-
	Significance Level									
	38.9%	11.3%	9.4%	32.6%	58.8%	60.4%	10.4%	16.5%	18.7%	26.2%
	t-values and One-sided Significance Level									
t-value	-.30	-1.27	-1.30	-.50	.23	.25	-	-	-	-
from RAND	38.2%	10.2%	9.7%	30.9%	40.9%	40.1%	-	-	-	-
t-value	-.27	-1.36	-1.60	-.63	.30	.30	-	-	-	-
implied by H&H	39.4%	8.7%	5.5%	26.4%	38.2%	38.2%	-	-	-	-

Table 10: Stratified Randomization Estimates of OLS Slope Sampling Distribution for Monthly Equal-Weighted NYSE Total Returns (CPI-Adjusted) from CRSP: 1926 - 1986

$$r_{K,t+K} = \alpha_K + \beta_K \; r_{K,t} + \epsilon_{K,t+K}$$

	Return Horizon (K: years)									
	1	2	3	4	5	6	7	8	9	10
	Historical OLS Slopes (β_K)									
	-.07	-.26	-.39	-.47	-.47	-.28	-.10	-.16	-.11	-.10
	Randomization Results									
mean	-.08	-.14	-.19	-.24	-.28	-.30	-.31	-.31	-.31	-.31
median	-.09	-.15	-.20	-.25	-.30	-.31	-.33	-.33	-.33	-.32
SD	.13	.16	.16	.17	.18	.19	.20	.22	.23	.25
	Fractiles									
1%	-.34	-.49	-.55	-.61	-.63	-.67	-.69	-.75	-.80	-.89
2.5%	-.31	-.44	-.50	-.55	-.59	-.62	-.64	-.68	-.74	-.78
5%	-.28	-.41	-.46	-.51	-.55	-.59	-.60	-.64	-.67	-.70
10%	-.24	-.35	-.40	-.46	-.51	-.53	-.54	-.56	-.60	-.62
20%	-.18	-.28	-.34	-.39	-.44	-.46	-.48	-.49	-.50	-.50
	Significance Level									
	55.2%	23.1%	11.2%	9.0%	14.0%	56.7%	84.2%	75.2%	81.0%	81.8%

Data Description

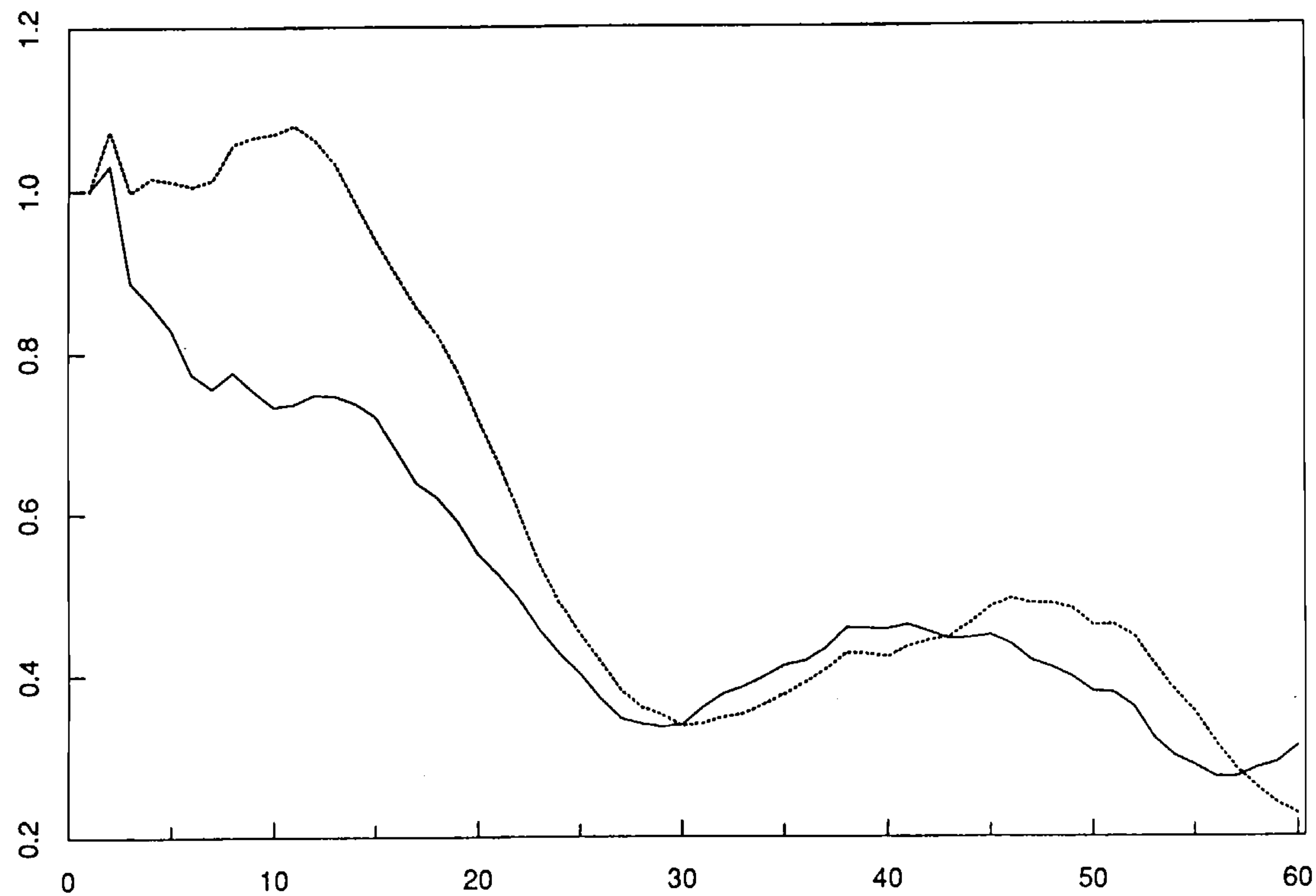
1. **Shiller's January average (1872 - 1986):** Annual total return for 1872 - 1986 period is derived from S&P 500 January average from Shiller [1981], revised and updated, and deflated by PPI. For example, continuously compounded real return (r_t) is derived from, $r_t \simeq ln(1 + r_t) \simeq ln(p_t/p_{t-1} + d_t/p_{t-1})$ where p_t = real January S&P 500 at time t and d_t = nominal dividend paid during time t divided by average PPI at t. Pre-1925 excess returns are derived using commercial paper rate (CP) reported in Gordon [1986] and one-month TB rate in Ibbotson Associates' Yearbook [1987] for post-1926 period.

2. **December closing (1871 - 1987):** Nominal total return (R_t) for December Closing S&P 500 is adopted from Ibbotson [1987] extended with Wilson Jones [1987]. It is adjusted for the inflation rate of CPI to derive real return (r_t), i.e., $r_t = (1 + R_t)/(1 + \pi_t) - 1$. For excess returns ($E_t$), $E_t = (1 + R_t)/(1 + TB_t) - 1$ where TB_t = annual T-bill rate extended with CP rate for pre-1925 period.

3. **Monthly NYSE returns (1926.1 - 1986.12):** Both equal- and value-weighted, including and excluding dividends, all NYSE stock returns are obtained from CRSP return files. To derive both real and excess returns we use the same formula as those used for December closing. CPI inflation rate and one month TB rate for 1926 - 1987 period are taken from the Ibbotson [1987]. All monthly returns are continuously compounded, e.g., $r_t = ln(1 + r_t)$.

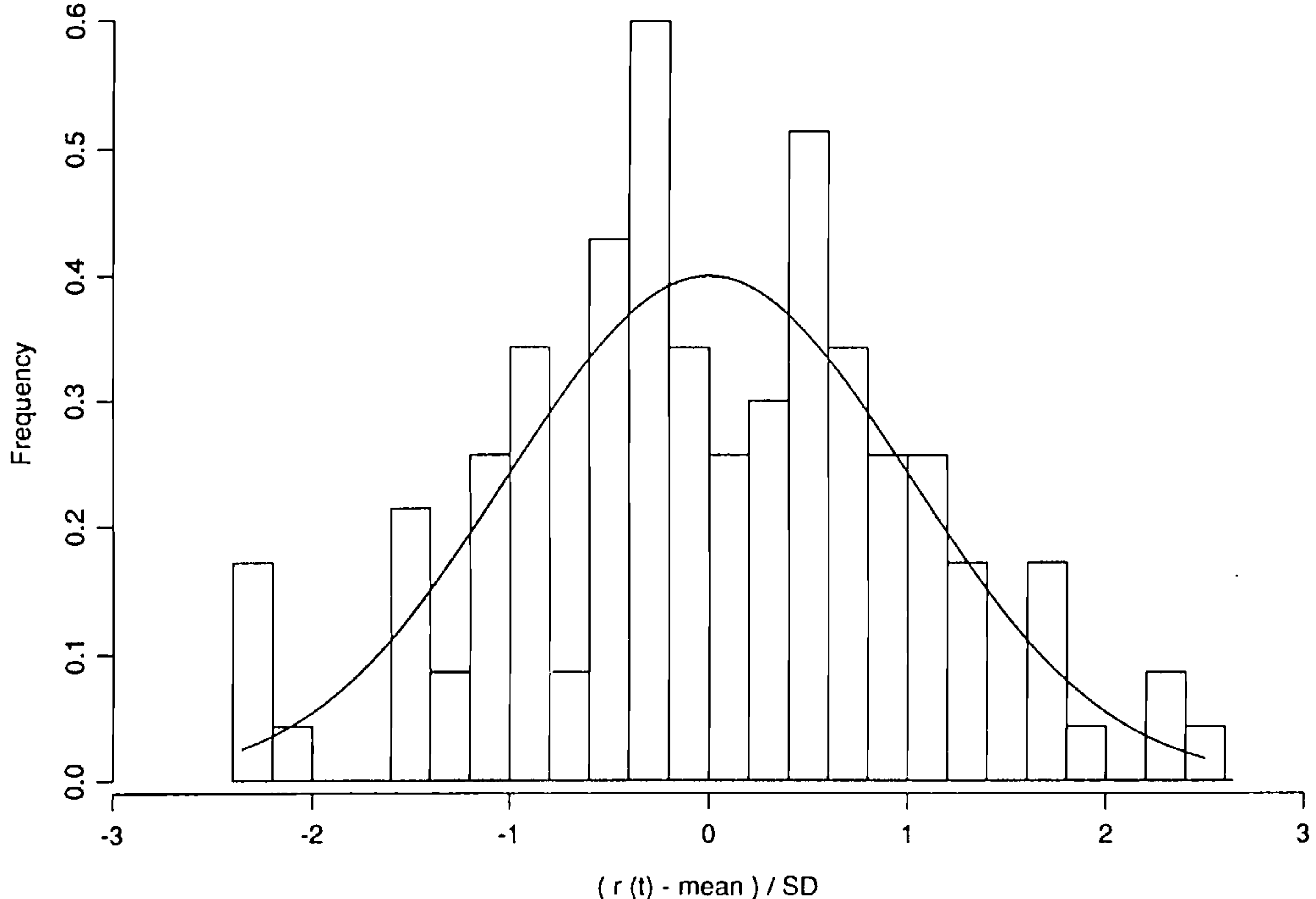
Note: Interested readers may request a copy of data sets examined here from the authors. It should be directly addressed to: Myung J. Kim, Department of Economics, DK-30, University of Washington, Seattle, WA 98195.

Figure 1: VR of Real Returns on the S&P Composite: 1872 - 1986



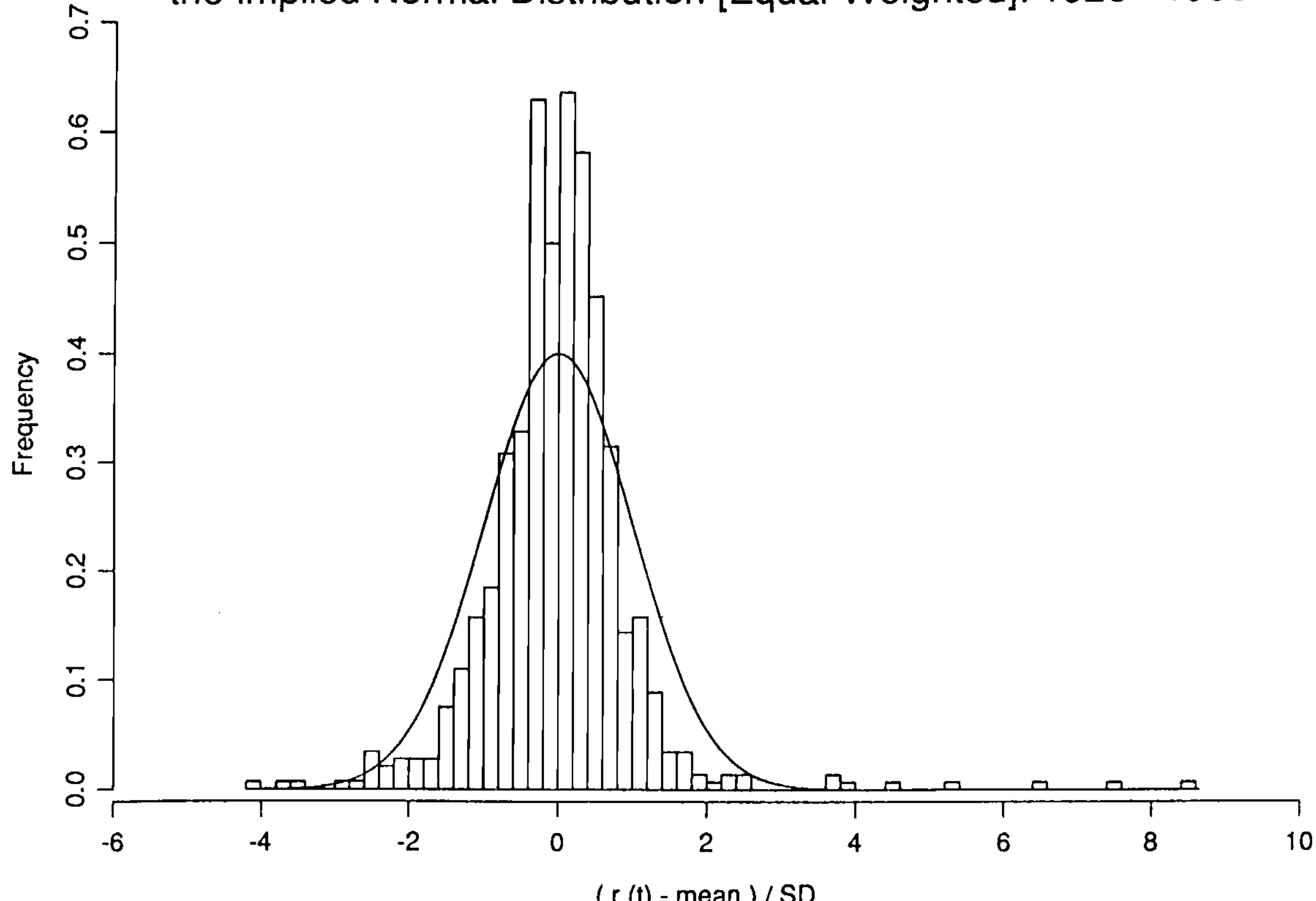
K: Years
Solid line is VR(K) estimated from December Closing (CPI-adjusted)

Figure 2: Frequency of Annual Real Returns (from December Closing) and the Implied Normal Distribution: 1871 - 1987



Superimposed is the corresponding Normal distribution

Figure 3: Frequency of Monthly Real Returns (from CRSP) and the Implied Normal Distribution [Equal-Weighted]: 1926 - 1986



Superimposed is the corresponding Normal distribution

Figure 4: Frequency of Monthly Real Returns (from CRSP) and the Implied Normal Distribution [Value-Weighted]: 1926 - 1986

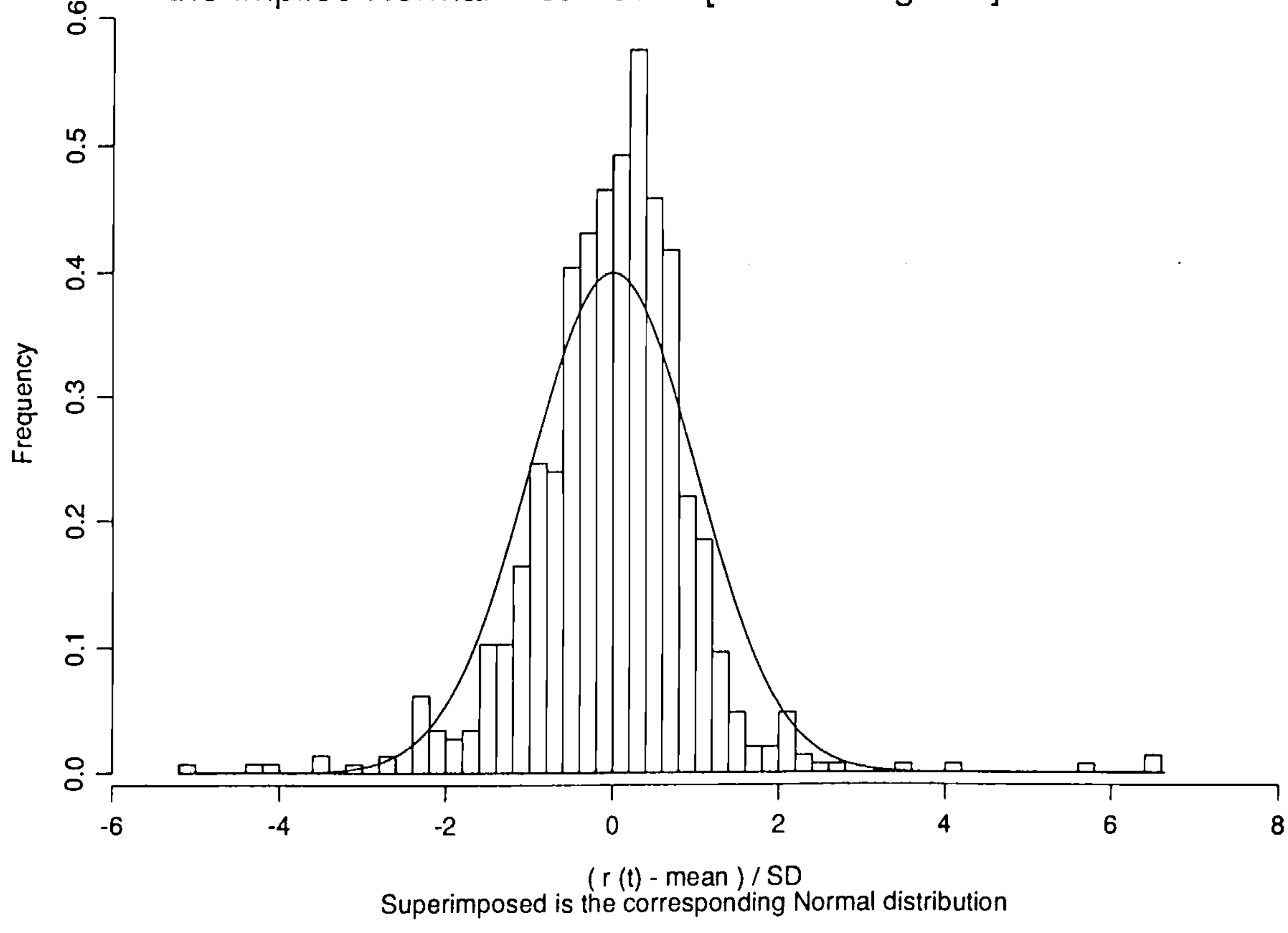
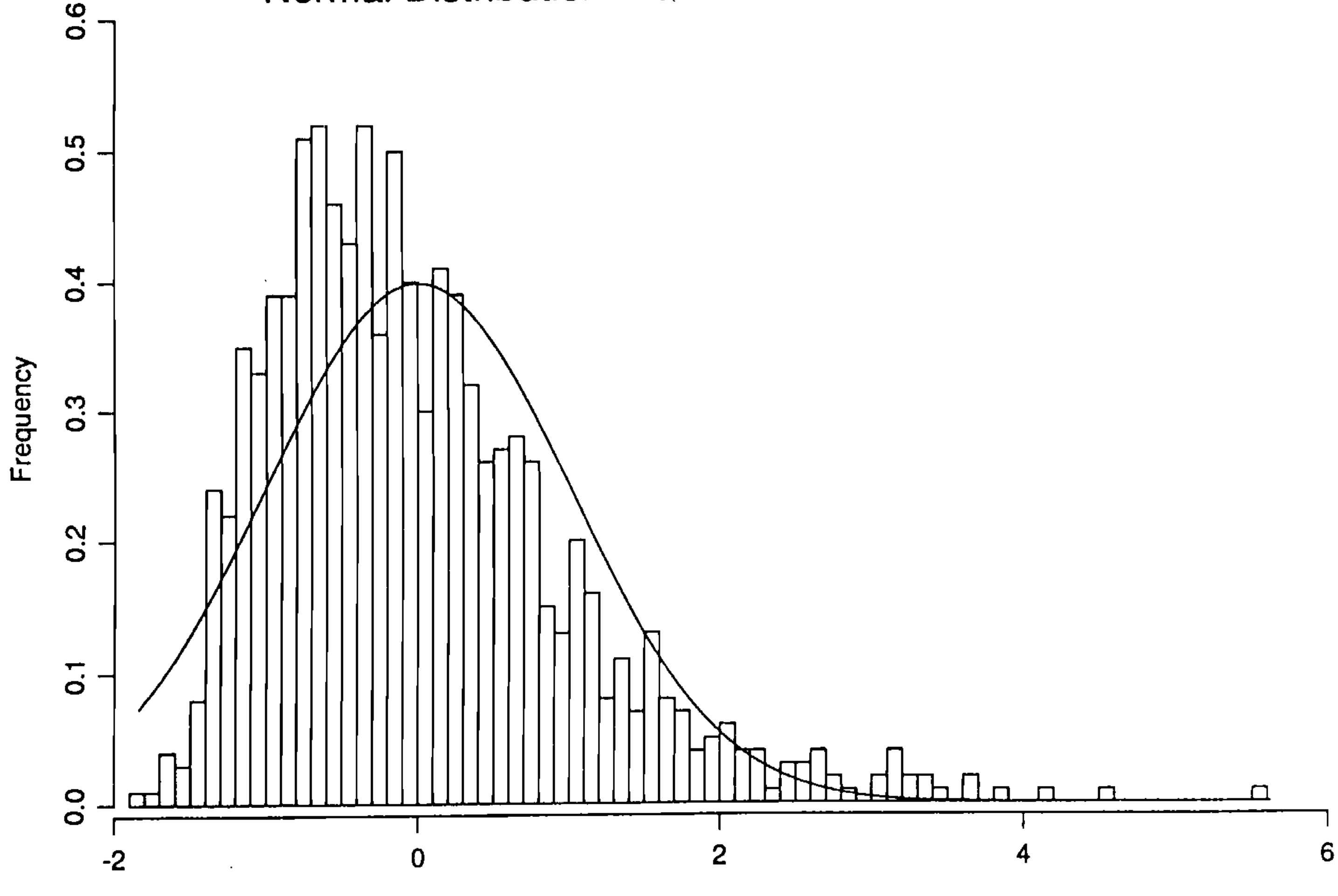


Figure 5: Sampling Distribution of VR(17) and the Corresponding Normal Distribution: Randomization for 1871 - 1987



Superimposed is the corresponding Normal distribution

Figure 6: Sampling Distribution of VR(19) and the Corresponding Normal Distribution: Stratified Randomization for 1871 - 1987

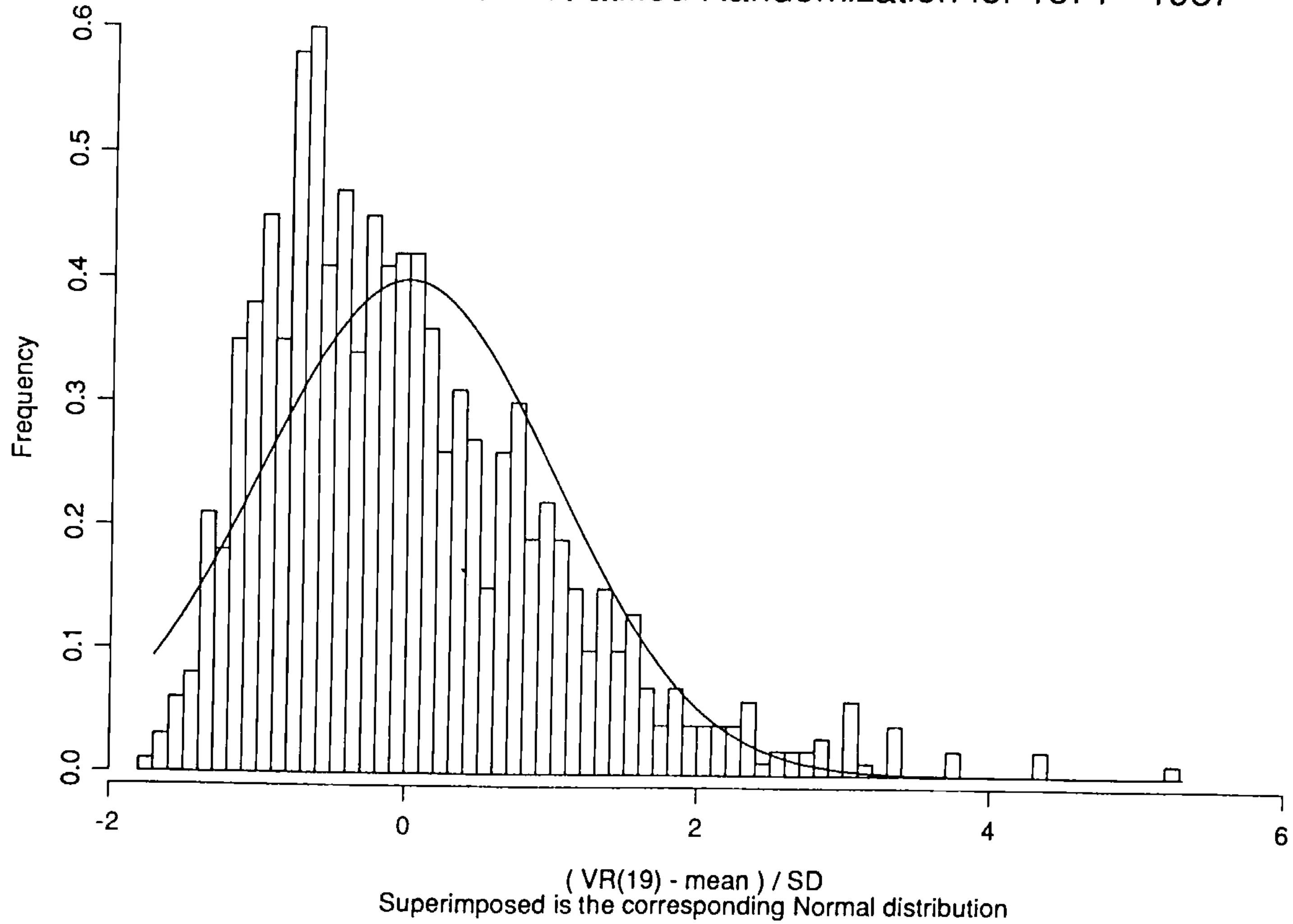
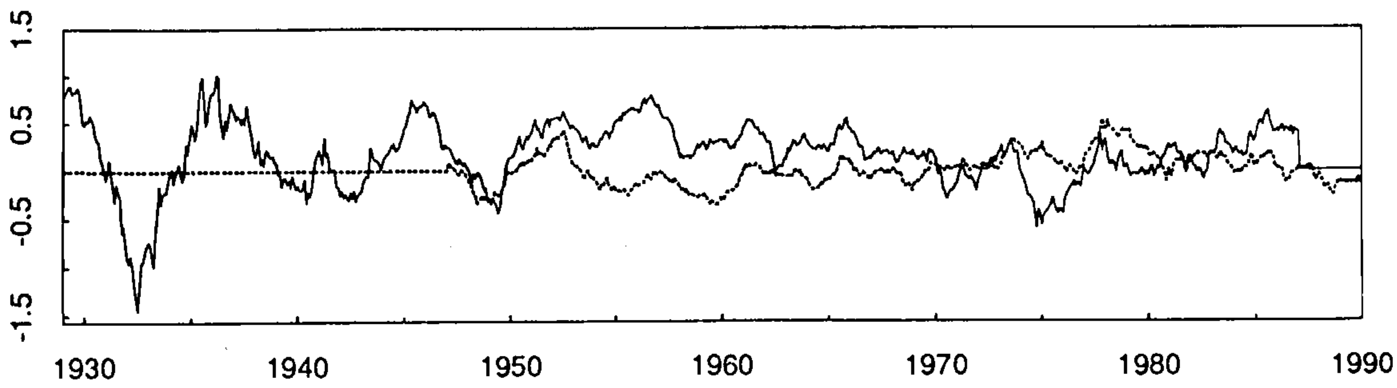
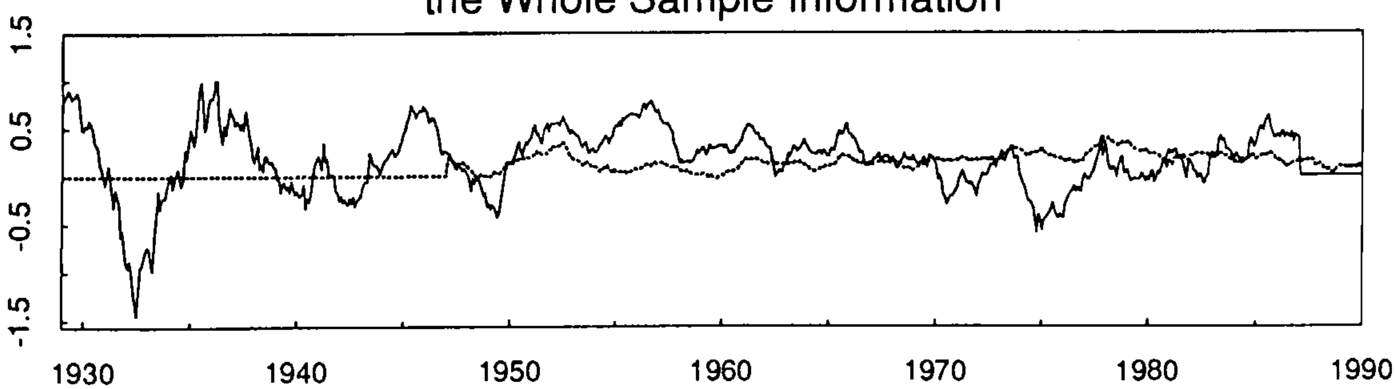


Figure 7-1: 3-Year Ahead Predictions: Based on the Pre-War Estimates



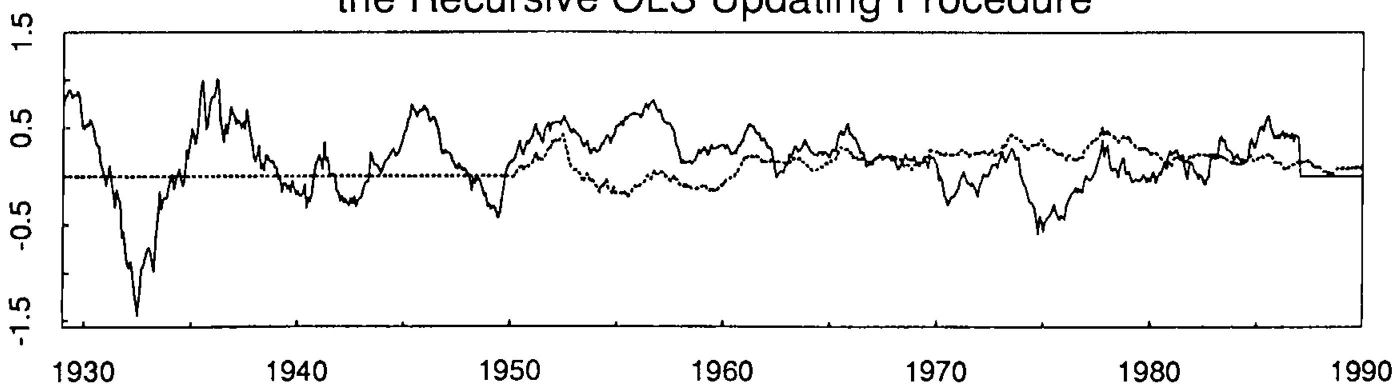
Data: Value-Weighted

Figure 7-2: 3-Year Ahead Predictions: Based on the Whole Sample Information



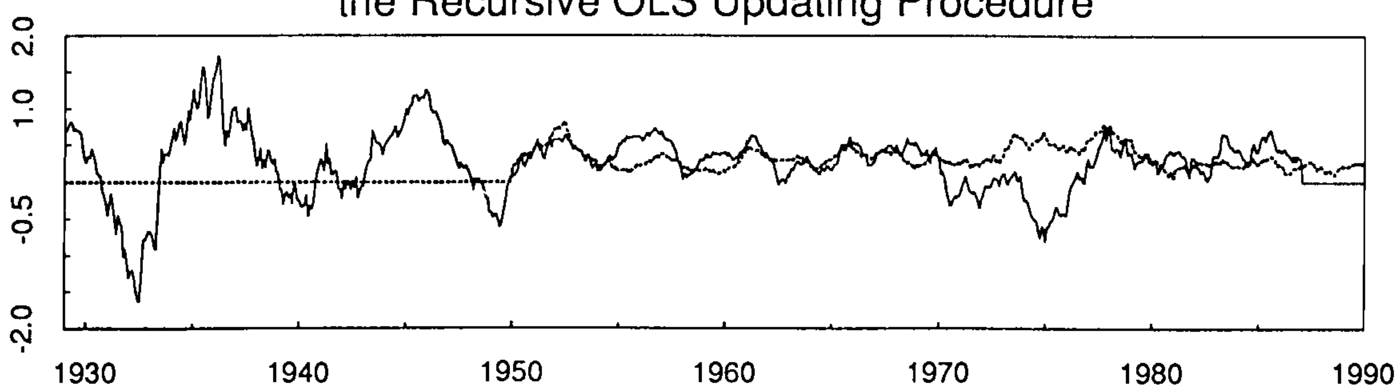
Data: Value-Weighted

Figure 8-1: 3-Year Ahead Predictions: Based on the Recursive OLS Updating Procedure



Data: Value-Weighted

Figure 8-2: 3-Year Ahead Predictions: Based on the Recursive OLS Updating Procedure



Data: Equal-Weighted