RESEARCH SUPPORT
UNIVERSITY OF MICHIGAN BUSINESS SCHOOL

AUGUST 1997

LONG HORIZON MEAN REVERSION IN THE STOCK MARKET: THE POSTWAR YEARS

WORKING PAPER #9712-20

JONATHAN PAUL CARMEL
UNIVERSITY OF MICHIGAN BUSINESS SCHOOL
AND
HEBREW UNIVERSITY

MARTIN R. YOUNG
UNIVERSITY OF MICHIGAN BUSINESS SCHOOL

LONG HORIZON MEAN REVERSION IN THE STOCK MARKET:

THE POSTWAR YEARS

Jonathan Paul Carmel
University of Michigan Business School
jpcarmel@umich.edu

and

The Jerusalem School of Business Administration
Hebrew University

Martin R. Young*
University of Michigan Business School
Ann Arbor, MI 48109-1234
Phone: 313-936-1332
FAX: 313-936-0274
myoung@umich.edu

AUGUST 1997

ABSTRACT

Several studies have found long-horizon mean reversion for small firm portfolios and the equally-weighted index for sample periods including the Depression. However, mean reversion has not been documented for sample periods which exclude the Depression. Similarly, mean reversion has not been established in any time period for the value-weighted index or for portfolios of large firms. These two characteristics of past findings suggested that mean reversion may have been a peculiarity of the Depression years and, particularly, of small firms during the Depression. We find highly significant evidence of mean reversion in the value weighted index for the postwar period. We consider 7 different predictors of returns, based on past returns, and find that the p-value against serial independence is .007 (two-sided), after correcting for multiple testing. The estimated model implies that 39.4% of a marginal return shock will eventually be reversed.

JEL Classification: G10, G14.

---

*Corresponding author.

1. Introduction

Is stock market mean reversion at annual and monthly frequencies an artifact of the prewar era? Is it restricted to small firms? In the past ten years, many studies have documented mean reversion in the equally-weighted stock market index in prewar data (Fama and French (1988), Kim, Nelson, and Startz (1991), Jegadeesh (1991)). Fama and French (1988) conclude that mean reversion remains detectable even when one extends the data to stretch from 1926 to the present. However, these findings have generated controversy.[1] But, one point of consensus in the literature is that to-date there has been little evidence of mean reversion at monthly or annual frequencies in the postwar sample (1947 to the present).[2] Another special feature of the evidence on mean reversion is that mean reversion has not been demonstrated for the value-weighted index or for large firm portfolios during either the whole CRSP sample or the postwar period.[3]

The inability of the literature to document mean reversion in the postwar sample provides support for the contention of Kim, Nelson, and Startz (1991) that any mean reversion which may exist in the CRSP time series of monthly returns is due to special characteristics of the Depression and may not be relevant for the current time period. The lack of evidence for large firm portfolios or for the value-weighted index during any time period suggests that mean reversion, when it exists, may due to the peculiarities of financial markets involving small firms and may not have relevance for much of the capital invested in the stock market. In this paper, we find highly significant evidence of mean reversion in the US value-weighted index during the postwar period. Thus mean reversion appears to be a current feature of the US market index, affecting large firms as well as small firms.

---

[1]See Richardson and Stock (1989), Jegadeesh (1991), Kim, Nelson, and Startz (1991), Richardson and Smith (1991), and Richardson (1993), Lamoureux and Zhou (1996), Bidarkota and McCulloch (1996).

[2]See Fama and French (1988), Richardson and Stock (1989), Jegadeesh (1991), Richardson and Smith (1991), Kim, Nelson, and Startz (1991), and Richardson (1993), Daniel (1994), Bidarkota and McCulloch (1996), Lamoureux and Zhou (1996).

[3]In Tables 2 and 4 of Fama and French (1988), none of the bias-adjusted slopes are significant for decile 9 or decile 10 portfolios (large firms) or the value-weighted index. Also see Tables II and III in Jegadeesh (1991). While not addressing mean reversion itself, Daniel and Torous (1995) find some evidence of predictability in the value-weighted index using all past lags as predictors. These results are discussd in more detail later in the paper.

Traditionally, mean reversion was thought to be an abnormal condition for financial market data. Thus the profession has taken some comfort from the fact that evidence of mean reversion had been restricted to the prewar, Depression years. However, since the wide adoption of ARCH methodology in the late 1980's, evidence of volatility clustering and predictability of volatility in the value-weighted index has been widely documented.[4] Thus, there is rigorous as well as casual evidence to support the view that market participants do not perceive the riskiness of the stock market to be the same in all months, but, instead, view some months to be high risk months and other months to be low risk. One would expect that the times of high risk should have high expected returns and low risk periods should have low expected returns.[5]

As many others have argued before, basic economic theory implies that shocks to risk will cause returns to have a mean reverting component: when a risk shock occurs that increases the riskiness of the market, investors will demand a higher expected return to remain in the market. Holding all else equal, this will cause the realized return on the market to be below its unconditional mean at the moment the risk shock occurs and the expected value of the future return to be above the unconditional mean. Similarly, a shock which reduces the riskiness of the market, will cause the realized return to be above the unconditional mean (all else equal) and the expectation of the upcoming future return to be below its long-term mean.

Given that the time series predictability of the riskiness of the returns on the value-weighted index has been amply demonstrated for the postwar sample, theory suggests that mean reve- ion should be present here as well. While there are other possible explanations for the inability to document mean reversion in the postwar value-weighted index, we approach this primarily as a problem of lack of power. The mean of stock market returns is notoriously difficult to estimate, due to the low signal to noise ratio of the returns process

[4]The survey article by Bollerslev, Chou, and Kroner (1992) cites many examples.

[5]This should occur unless the equilibrium risk premium on market risk or some other component of the discount function varies in such a way as to completely offset the effect coming from time-state varying risk.

(see, e.g., Merton (1980)). A standard model of stock market returns is that the return at time t can be modeled as the sum of two independent components: $r_t = \mu_t + \varepsilon_t$ where μ_t is potentially serially correlated and ε_t is mean zero and serially uncorrelated. There is controversy about the nature of the μ_t component. Some would argue that it is a constant (at least in the postwar sample). Others would argue that it is mean reverting. But there is general agreement that the variance of the serially uncorrelated process ε_t is much higher than the variance (if any) of the μ_t process. The high variance of the serially uncorrelated ε_t process makes the unconditional mean of the return time series r_t hard to estimate and certainly makes it difficult to test subtle hypotheses about the nature of the μ_t process. Summers (1986) and Poterba and Summers (1988) argue that, in particular, it may be difficult to detect slowly mean reverting processes using high frequency autoregressions.

Other evidence that power may be an issue is that several authors have demonstrated the ability to predict the conditional mean of the value-weighted index in the postwar period based on other instruments such as dividend yields (Fama and French (1988), Hodrick (1992)) and industrial production (Daniel and Torous (1995)). Thus it appears that dividend yields and industrial production contain additional information that allow one to forecast the conditional mean of stock market return. Therefore, one might expect that, with more powerful inference techniques, one might be able to find univariate predictability in the returns on the value-weighted index. In fact, Daniel and Torous find weak evidence of univariate predictability in the value-weighted index in the postwar period. Daniel and Torous calculate smoothed autocorrelations for all lags for the value-weighted index and test the joint hypothesis that $\rho_1 = \rho_2 = \dots = \rho_N = 0$ where ρ_n is the nth order smoothed autocorrelation. They report three different levels of smoothing, and can reject the null hypothesis of serial independence for one of these levels at the 5% level.[6] Rejection of the joint hypothesis $\rho_1 = \rho_2 = \dots = \rho_N = 0$ does not give specific information about the sign, magnitude, or duration of the effect of shocks to the μ_t or r_t process; so while the result

[6]For the equally-weighted index, Daniel and Torous reject the serial independence hypothesis for all three levels of smoothing.

implies predictability of the returns process, it does not necessarily signify the presence of long-horizon mean reversion.

We increase the power of previous tests by employing a GARCH error specification to correct for the time-varying volatility of the return series. Given the overwhelming evidence of variance clustering in the value-weighted index and the need to increase the power of inference techniques in investigations of stock market mean reversion, it seems natural to incorporate GARCH techniques directly into the estimation procedure.

We find highly significant evidence of mean reversion in the value weighted index for the postwar period. In particular, we find that the average of the previous 24 monthly returns has a t-statistic of -3.26 when used to predict the return in the upcoming month. The mean reversion present at this frequency has the effect of reversing, eventually, 39.3% of a marginal return shock. We consider seven averaging lengths (12, 24, 36, 48, 60, 72, and 84 months) and recognize that the estimation procedure has searched over seven possible averaging lengths. To obtain the joint significance of the -3.26 t-statistic, we derive the joint distribution of the seven test statistics and find that the corrected p-value is 0.007 (two-sided) which remains highly significant. We also use the joint distribution of the test statistics to transform the original seven estimates into seven independent tests. We find that the joint significance of these seven independent tests is 4.33×10^{-5}. Finally, we employ a χ^2 test similar to that in Richardson and Smith (1991) to assess the joint significance of the original seven test statistics. This χ^2 test reports a p-value of 3.57×10^{-5}. We check the robustness of our findings by examining annual US data during the sample period (1947-1995) using OLS. We then repeat the OLS analysis using annual returns for the US prior to the CRSP sample (1885-1925) from the Schwert (1990) data set and for several foreign stock markets from 1951-1995, and find that mean reversion at a two-year frequency exists in these series as well, thus offering corroboration of the initial findings.

2. The Statistical Methodology

2.1 The Model

We investigate the predictability of returns by regressing the current return on a regressor constructed by averaging past returns.[7] Define $\bar{r}_{t-1}^{(k)} = \frac{1}{k}\sum_{j=1}^{k} r_{t-j}$ where k is a number of past returns. We estimate the model

$$r_t = \beta_0 + \beta\,\bar{r}_{t-1}^{(k)} + e_t, \tag{1}$$

where e_t follows a GARCH process and k is a parameter of the model which is estimated by maximum likelihood from a set of candidate values. Equation (1) expresses the conditional mean of r_t as an affine function of $\bar{r}_{t-1}^{(k)}$. It is well-known that it is difficult to estimate the mean of the value-weighted index from the univariate time series (See Merton (1980), Hodrick (1992)). Therefore only powerful estimation procedures have a chance at detecting significant evidence of mean reversion. (See Summers (1986) and Poterba and Summers (1988)). This consideration leads us to impose a fairly sparse parameterization, since any highly general specification will tend to have low power to detect the phenomenon in question. One of the features of estimating model (1) is that it has only three free parameters in the mean equation and yet cannot bias point estimates or standard errors toward falsely rejecting a null of serial independence. The model $r_t = \beta_0 + \beta\,\bar{r}_{t-1}^{(k)} + e_t$ can be rewritten as:

$$r_t = \beta_0 + \sum_{j=1}^{\infty} \beta_j r_{t-j} + e_t$$

with the constraints $\beta_1 = \beta_2 = \ldots = \beta_k$ and $\beta_{k+1} = \beta_{k+2} = \ldots = 0$. These two sets of constraints cannot bias one toward finding predictability if the underlying data comes from a serially independent process. Similarly, these constraints cannot induce serially correlated model errors. This is in marked contrast to procedures which use averaged returns on the lefthand side of the regression equation and employ rolling overlapping regression windows. Regressions which use overlapping lefthand side variables require one to correct for induced biases in both point estimates and standard errors.

[7] Jegadeesh (1991) and Hodrick (1992) also use averages of past returns to predict the current return.

Further, retaining unaveraged returns as dependent variables has been shown to increase power against interesting alternative hypotheses. Jegadeesh (1991) shows that, asymptotically, one maximizes power against an AR(1) alternative by averaging to form the right hand side variable while leaving the lefthand side variable unaveraged. Daniel (1994) also shows that using averaged variables to form the righthand side regressor in a time series prediction equation can improve power. However, Daniel finds that, for the class of local alternatives, the same increase in power can be achieved asymptotically whether one averages to form the dependent variable or the independent variable. Finally, Hodrick (1992) shows that restricting averaging to the righthand side has superior small sample properties. This remains true even after one uses various procedures to correct for the serial correlation induced when one also averages to form the lefthand side variable.

Methodologically, our paper is closest to Jegadeesh (1991), which also investigates mean reversion using a regression model with aggregated returns as the predictor variable and disaggregated returns as the dependent variable. However, in the spirit of Richardson and Smith (1991), we derive the joint distribution of the univariate test statistics in any case in which our estimation procedure searches over a set of possible averaging windows. In contrast to Jegadeesh (1991) and Richardson and Smith (1991), we incorporate a GARCH error specification when estimating parameter values. The addition of the GARCH specification is particularly important, given the need to take reasonable measures to enhance power in order to understand the mean behavior of the time series of stock returns. (See Section 3.5.)

2.2 *Estimation of the Length of the Averaging Window k*

To illustrate the value of using averaged returns as the RHS predictor, it is useful to view the returns process as being the sum of two independent components:

$$r_t = \mu_t + \varepsilon_t$$

where μ_t is potentially serially correlated and ε_t is mean zero and serially uncorrelated. Since any predictability in the r_t process must by definition come from the μ_t process, one would have the greatest power to detect mean reversion if one could observe μ_t and directly regress μ_t on past values of the μ_t process. However, the μ_t series is unobservable. Instead one can only see r_t which can be viewed as an observation of μ_t with error. By averaging past returns to form the righthand side variable in Equation (1), we reduce the influence of the ε_t process on the predictor (RHS) variable. However, averaging also smears the variation in the μ_t process, tending to decrease the power of the righthand side predictor. The optimal averaging length k^* must balance the power gain that comes from decreasing the impact of the ε_t series on the RHS predictor against the power loss that comes from losing information about the exact arrival time of each μ_t. Our estimation procedure chooses the k that maximizes the likelihood function of Equation (1).

2.3 Assessing Significance

In the context of a univariate regression model such as $r_t = \beta_0 + \beta\, \bar{r}_{t-1}^{(k)} + e_t$, the k that maximizes the model's likelihood function will be equivalent to the k that yields the largest t-statistic (in absolute value) for the estimated β. Thus our procedure for estimating k is mathematically equivalent to searching for the largest magnitude t-statistic over N possible righthand side regressors where N is the number of candidate values for k.

To assess significance, we derive the asymptotic joint distribution of the N t-statistics under the null that the data are identically and independently distributed. We then correct for the fact that the t-statistic on the estimated β is equivalent to the largest magnitude t-statistic sampled from a set of N correlated t-statistics.

Let $\hat{\beta}(k')$ be the OLS estimate of β in Equation (1) for k set equal to k'. Also let $t(k')$ be the t-statistic associated with $\hat{\beta}(k')$. Proposition 1 derives the asymptotic distributions of $\hat{\beta}(k)$ and $t(k)$ under the null. All proofs are in the appendix.

Proposition 1. Under the null that returns are iid, the asymptotic joint distribution of $\hat{\beta}(k)$ satisfies

$$T\,\mathrm{Var}(\hat{\beta}(k)) = k \qquad \rho(\hat{\beta}(k_1), \hat{\beta}(k_2)) = \frac{k_1}{\sqrt{k_1 k_2}}$$

and the asymptotic joint distribution of $t(k)$ satisfies

$$\mathrm{Var}(t(k)) = 1 \qquad \rho(t(k_1), t(k_2)) = \frac{k_1}{\sqrt{k_1 k_2}}$$

where $k_1 \leq k_2$ and ρ denotes correlation.[8]

Richardson and Smith (1991) derive results analogous to Proposition 1 for analyses involving overlapping LHS observations. They find that the asymptotic joint covariance of the estimated slope coefficients $\hat{\beta}(k_1)$, $\hat{\beta}(k_2)$, ... under the null does not depend on any sample estimates but instead is only a function of lag lengths k_1, k_2,..., k_N and the sample size T. We find our estimators have this same property. In particular, the standard error of individual estimated slope coefficient may be calculated without reference to any sample estimates. One can assess significance by dividing the estimated slope coefficient $\hat{\beta}(k)$ by $\sqrt{k/T}$ which is the standard error of $\hat{\beta}(k)$ under the null as derived in Proposition 1.[9]

3. Estimation and Mean Reversion Evidence

3.1 The Data, Averaging Window, and GARCH Specification

We use monthly data from January of 1947 to December of 1995.[10] For each month, r_t is set equal to the excess return of the value-weighted index where data for the value-weighted index and risk-free rates are taken from the CRSP data files. We estimate k from

[8] The main contribution of this proposition is in providing the correlations of the point estimates and their t-statistics under the null. These correlations are needed in order to properly correct for multiple inference. Jegadeesh (1991) derives a generalized formula for the univariate variance of the point estimates under an iid null; our univariate variance result is a special case of his result.

[9] Proposition 1 assumes that the data are homoskedastic. However, there is significant evidence of heteroskedasticity in the time series of stock returns. We correct for this by incorporating a GARCH error specification and estimating parameters by maximum likelihood estimation, rather than OLS. Table 4 (discussed later in the paper) shows that the empirically observed standard errors for the maximum likelihood estimates closely match the values predicted by Proposition 1.

[10] The actual data set starts in January 1926. As necessary, we use data prior to January of 1947 to form the righthand side variable $\bar{r}_{t-1}^{(k)}$.

the data by considering 12, 24, 36, 48, 60, 72, and 84 months as possible values of k. Following Attanasio (1991) and Bodurtha and Mark (1991), we find 3rd order terms are valuable in describing the heteroskedasticity in monthly returns and use an IGARCH(3,3) specification for the error term.[11] In particular, the 3-month squared residual is especially important for predicting volatility in model (1) and has a t-statistic of 2.74. Bollerslev, Chou, and Kroner (1992) suggest that 3rd-order terms may be important in explaining the volatility of monthly stock returns due to possible clustering effects in the quarterly announcements of dividends and earnings. We find that estimates of β in Equation (1) are stable with respect to changes in the particular GARCH specification chose for e_t.

3.2 Main Results

For the January 1947 to December 1995 period, Table 1 reports that the optimal value for k is 24 months, the estimated slope coefficient on $\bar{r}_{t-1}^{(24)}$ is -.66, and the t-statistic is -3.26 which, in the context of estimating a single regression, corresponds to a p-value of 0.0011.[12] The probability under the null that at least one of seven t-ratios will have a magnitude of 3.26 or greater, given that the t-ratios have the correlation structure specified in Proposition 1, is 0.007; thus the finding remains highly significant after correcting for multiple inference.[13] By construction, the seven candidate righthand side regressors are highly correlated (see Table 2). Therefore, the seven test statistics are highly correlated, and we have many fewer than seven independent tests being performed. However, one would have to snoop over 45 <u>independent</u> tests in order to have a 5% chance of arriving at a t-statistic of -3.26 due to random sampling error.[14]

---

[11]The IGARCH(3,3) specification states that the variance of e_t, which will be denoted by h_t, follows the process $h_t = \alpha_0 + \alpha_1 h_{t-1} + \alpha_2 h_{t-2} + \alpha_3 h_{t-3} + \phi_1 e_{t-1}^2 + \phi_2 e_{t-2}^2 + \phi_3 e_{t-3}^2$, where α_0, α_1, α_2, α_3, ϕ_1, ϕ_2, ϕ_3 are parameters to be estimated, subject to the restriction that $\alpha_1 + \alpha_2 + \alpha_3 + \phi_1 + \phi_2 + \phi_3 = 1$.

[12]All significance levels reported in this paper are with respect to 2-sided tests, except for χ^2 tests which are 1-sided.

[13]-3.26 is the theoretical t-ratio obtained by dividing the estimated slope coefficient (-.66) by the asymptotic standard deviation of the slope coefficient under the null as derived in Proposition 1 (.20). Alternatively, if one computes the empirical t-ratio using the maximum likelihood standard error, the value of t is -3.30 which corresponds to a joint p-value of 0.006.

[14] n equal to 45.6 satisfies the equation $0.95 = (1 - 0.001124)^n$.

Proposition 2 enables us to transform the original seven regression results into seven independent tests.

Proposition 2. Let κ be a positive integer. Consider a set of equally spaced lags k_1, k_2,...,k_N such that $k_n = n\kappa$.

$$\text{Let } \hat{\gamma}(k_n) = \hat{\beta}(k_n) - \hat{\beta}(k_{n-1}) \qquad \text{for } 1 < n \leq N$$
$$= \hat{\beta}(k_n) \qquad \text{for } n = 1.$$

Under the null that returns are iid, the asymptotic joint distribution of $\hat{\gamma}(k_n)$ satisfies:

$$T\ \mathrm{Var}(\hat{\gamma}(k_n)) = \kappa \qquad \rho(\hat{\gamma}(k_m), \hat{\gamma}(k_n)) = 0 \text{ for } m \neq n.$$

From Proposition 2, the estimates $\hat{\beta}(12)$, $\hat{\beta}(24) - \hat{\beta}(12)$, $\hat{\beta}(36) - \hat{\beta}(24)$, $\hat{\beta}(48) - \hat{\beta}(36)$, $\hat{\beta}(60) - \hat{\beta}(48)$, $\hat{\beta}(72) - \hat{\beta}(60)$, and $\hat{\beta}(84) - \hat{\beta}(72)$ are all uncorrelated and each has a variance of 12/588 = 1/49. Therefore the t-statistic for $\hat{\beta}(24) - \hat{\beta}(12) = (-0.658 - -0.012)/(1/7) = -4.52$. (See Table 3.) The p-value associated with -4.52 is 6.19×10^{-6}. However, -4.52 is the largest t-ratio from among seven tests. The overall p-value, correcting for the fact that -4.52 represents the most extreme of seven independent t-statistics, is 4.33×10^{-5}.[15] In order to have a 5% chance of obtaining a t-statistic of this magnitude under the null, one would have to snoop over more than 8000 independent tests.[16]

The final test we use to assess the joint significance of the seven regressions estimated by our procedure is a χ^2 test, as used in Richardson and Smith (1991). Let b be the (Nx1) vector of estimated slope coefficients ($\hat{\beta}(k_1)$,... $\hat{\beta}(k_N)$), and let V be the asymptotic covariance matrix of b, as specified in Proposition 1. Then $bV^{-1}b$ asymptotically has a χ^2 distribution with N degrees of freedom. In our case, the value of this test statistic is 32.3 which corresponds to a p-value of 3.57×10^{-5}.

While the joint significance of the results from Table 1 are highly significant, the joint significance of both (i) the test based on the $\hat{\gamma}(k)$ values and (ii) the chi-square test are even greater; the joint significance of the first is 0.007 compared to 4.33×10^{-5} and 3.57×10^{-5}

[15] $1 - [1 - 6.19 \times 10^{-6}]^7 = 4.33 \times 10^{-5}$.

[16] n equal to 8286 satisfies the equation $0.95 = (1 - 6.19 \times 10^{-6})^n$.

for the other two tests. The difference derives from the fact that the first test statistic, the largest magnitude t-statistic, ignores the information about the null hypothesis contained in $\hat{\beta}(k)$ for values of k other than 24.[17] The other two test statistics use the information in point estimates other than $\hat{\beta}(24)$ in order to evaluate the probability of the null. In our case, these other point estimates, especially $\hat{\beta}(12)$, contain information that leads the other two tests to assess that the probability that the null hypothesis is true is substantially lower than that found in the initial test. In particular, $\hat{\beta}(24)$ is very high while $\hat{\beta}(12)$ is not; these two results imply that there is a very large amount of sample predictability coming from monthly lags 13 through 24, much more than can be explained by chance.[18]

The joint significance levels were computed based on Propositions 1 and 2, which were derived for an OLS regression estimator under the assumption of homoskedacity. In fact, the data are markedly heteroskedastic, but we have corrected for this through incorporating a GARCH error specification. Table 4 compares the theoretical t-statistics, attained by dividing the point estimates $\hat{\beta}(k)$ by the theoretical standard error derived in Proposition 1, to the empirical t-statistics calculated by dividing the point estimates $\hat{\beta}(k)$ by the observed maximum likelihood standard errors. The table shows that the theoretical t-statistics match most of the empirical t-statistics quite closely. This provides us with some confidence in applying Propositions 1 and 2 to assess the significance of the maximum likelihood estimates.

3.3 Magnitude and Nature of Reversals

Since the estimated slope coefficient from Equation (1) is negative and has magnitude less than 1, a shock to the return in month 0 causes oscillations in subsequent returns that dampen over time. Figure 1 graphs the shifts to subsequent returns that occur in response

[17]More precisely, the only information the test uses about these estimates is that their univariate t-statistics are lower in magnitude than the t-statistic corresponding to k=24.

[18]The relationship between the three joint test statistics considered in this paper can be viewed as follows: The first test calculates the significance of the largest magnitude t-statistic from a set of N correlated t-statistics. The second test, involving the $\hat{\gamma}$ estimates, transforms the initial t-statistics into N independent t-statistics and calculates the significance of the largest magnitude t-statistic from a set of N independent t-statistics. The chi-square test calculates the probability of the sum of the squares of the N transformed, independent t-statistics.

to a one-time shock in month 0, along with the cumulative percentage reversal of the initial shock, for the parameters values set equal to their maximum likelihood estimates. The oscillating pattern begins with negative returns for months 1 to 24 and then switches to positive returns for months 25 to 42 and continues to repeat in this manner, always with diminishing amplitude.[19]

One measure of the magnitude of the mean reversion in the data is the percentage of a one-time shock to returns which is reversed in the subsequent cumulative return. For infinitesimal shocks, the percentage of a time zero shock which is reversed in the cumulative return from time 1 to infinity can be calculated from the formula $\beta/(1+r_f+\beta_o-\beta)$ where β_o and β are the coefficients in Equation (1).[20] Setting parameter values equal to those estimated in Table 1 and the risk-free rate equal to its sample average, the formula implies that 39.3% of a return shock will eventually be reversed. From Figure 1, we see that while a shock at time 0 has the greatest effect on monthly returns at month 1, the cumulative effect peaks at 24 months. At 24 months, 48.3% of the initial shock has been reversed, compared to a final effect of only 39.3%. (Also see Table 5.)

3.4 Results for Yearly Data

The monthly results are confirmed by regression analysis on annual data. Here we restrict attention to the two year (24 month) averaging window estimated in the previous subsection and estimate

$$r_t = \beta_o + \beta \bar{r}_{t-1}^{(2)} + e_t \tag{2}$$

where r_t is now the excess return[21] of the value-weighted index in year t and $\bar{r}_{t-1}^{(2)}$ is the average of the excess return over the previous two years. The results are presented in Table

---

[19]To understand the cause of this pattern, note that the value of $\bar{r}_{t-1}^{(24)}$ decreases from months 1 to 24 as an increasing number of negative returns enter into the average $\bar{r}_{t-1}^{(24)}$. Thus the impact of the time-zero shock lessens from months 1 to 24. By month 25, the previous 24 returns have all been negative when parameter values are set equal to their maximum likelihood estimates. This means that $\bar{r}_{t-1}^{(24)}$ is negative at month 25 which, in turn, implies that month 25 return will be positive. Also see Table 5.

[20]This formula assumes the risk-free rate is constant and is derived in Appendix B.

[21]The excess return in year t was formed by compounding the monthly excess returns.

6. The point estimate for β is -.46 and the t-statistic is -2.21 which has a p-value of 0.03.[22] Twenty-three percent of any shock to annual returns is expected to be reversed in the following year,[23] but, from the R^2 of the regression, this explains only 9.4% of the total variation in annual returns. Thus if mean reversion is due to market over-reaction, individuals would still bear a large amount of market risk in any trading strategy that attempted to exploit this for profit. There are no significant GARCH effects in the annual data; therefore, this analysis is conducted using OLS.

The difference in the estimated mean reversion coefficients from the annual OLS regression (-.46) and the monthly regression with GARCH error specification (-.66) appears to be due to outliers in the yearly data occurring in 1974 and 1975. Analyzing the yearly data with an estimator which is robust to outliers, the least absolute deviations estimator (Bassett and Koenker (1978)), leads to an estimated slope coefficient of -.56 and a t-ratio of -2.18 (p-value = 0.035). Figure 2 plots annual excess returns against the average excess return over the previous two years. The scatter plot has a visible downward slope, illustrating the mean reversion effect. The two prominent outliers in the scatter plot correspond to the returns in years 1974 and 1975. An OLS analysis of Equation (2) which includes a dummy variable for the 1974-1975 period leads to an estimated slope coefficient on $\bar{r}_{t-1}^{(2)}$ of -.70, quite close to the monthly finding of -.66, and the t-statistic is -3.06; the t-statistic on the dummy variable is -2.20.

3.5 Importance of Using GARCH

Variance clustering is a well-documented feature of monthly stock market returns (see, e.g., Bollerslev, Chou, and Kroner (1992) for a survey of findings concerning time-varying volatility in stock returns). Some months contain a large amount of information about the mean process relative to others. Modeling errors as a GARCH process increases the power of the estimation procedure by appropriate up-weighting of informative

[22]The White (1980) t-statistic is -2.48 which has a p-value of 0.013.

[23]A positive unit shock in year zero will shift $\bar{r}_{t-1}^{(2)}$ up by one-half for t=1.

observations. We find that this increase in power is necessary in order to detect the evidence of mean reversion in the monthly data. The magnitude of the t-statistic on the slope coefficient in $r_t = \beta_0 + \beta\, \bar{r}_{t-1}^{(24)} + e_t$ drops to -1.67 when estimated on monthly data from January 1947 to December of 1995 using OLS, as compared to a t-statistic of -3.3 when an IGARCH(3,3) specification is used. One might be concerned that the mean reversion results have been induced by choosing a special GARCH specification. However, the OLS results on annual data make it clear that this is not the case. Also, the results are robust across GARCH specifications in the monthly data, with t-statistics beyond -2.3 obtained for a wide range of specifications.

3.6 Subsample Analysis

Equation (2) was also estimated on annual data for two subsamples: 1947 to 1971 and 1972 to 1995. For both subsamples, the estimated coefficient on $\bar{r}_{t-1}^{(2)}$ is negative (-.37 for the first subsample and -.76 for the second subsample). The subsample analysis contains no evidence contrary to a hypothesis of mean reversion. However, the number of observations in each subsample drops to 25 and 24, making it difficult to obtain statistically significant results. The t-statistics on these estimates are -1.22 for the first subsample and -2.51 for the second.

3.7 Out of Sample Analysis: Foreign Stock Markets 1951-1995, US 1885-1925

Equation (2) was also estimated for several countries listed on the DRI Basic Economics database. For these series, dividend and risk-free rate information were not available. Consequently, real stock index values were computed by dividing the nominal stock index by the country's consumer price index. The series r_t was then formed by taking the percentage change of the value of this real stock index at time t.[24] The data sample runs from 1951 to 1995. Table 7 contains the results. The estimated slope coefficient is negative

[24]The omission of dividends from the returns series can cause results to be biased. One stylized fact about dividend yields is positive autocorrelation: dividend smoothing. To the extent that percentage capital gains and dividend yields are not negatively correlated in time series, this will cause some bias toward finding mean reversion in the percentage capital gain series, even if there is none in the unobserved return series. On the other hand, strong negative correlation between percentage capital gains and dividend yields can cause bias against finding mean reversion in the price series.

for all four countries: Canada, France, Japan, and the United Kingdom. The estimated slope coefficient is significant for the U.K. (t-statistic -2.13) and nearly significant for Canada (t-statistic -1.87). Figure 3 plots the returns on the British index against the average lagged return from the past two years for 1952 to 1995. The plot reveals a clear downward slope. The observations corresponding to the years 1974 and 1975 appear as outliers with a similar pattern to that seen in the US data.

Given that the time series properties of US stock returns during the Depression have been intensively studied, we use the return series described in Schwert (1990) from 1885 to 1925 for an out of sample test.[25] This return series includes estimates of dividends. Each observation is the real annual return on the index which was computed as the difference between the annual return on the Schwert index and the percentage change in the price index for the corresponding year; price index data is from McCusk (1992). Equation (2) was estimated on annual US data for this period. Table 7 reports the analysis; the point estimate for β is -.54 and the t-statistic is -2.38, corresponding to a p-value of 0.02.

3.8 Large Firms

To-date evidence supporting mean reversion has been lacking for large firm portfolios, raising questions as to whether mean reversion is a broad macroeconomic phenomenon or a special feature of small firms. The current paper's results on the value-weighted index suggest that mean reversion is present, not just in the returns of small firms, but in the returns of large firms, as well. To verify this, we formed an equally-weighted portfolio consisting of the largest quintile of firms based on the size designation in the CRSP files (market value of equity). Equation (1) was then re-estimated for monthly returns and joint test statistics across all seven lags were calculated. Equation (2) was re-estimated as well on annual data using OLS. Table 8 finds highly significant evidence of mean reversion at the

[25]The returns data assembled in Schwert (1990) from 1885 to 1925 uses the Dow Jones Index to measure capital gains and the dividend yield on the Cowles portfolio to measure dividends. However, Dow Jones data does not exist prior to 1885. So at this point the Schwert time series switches to a different data source, causing a discontinuity in the time series properties of the data. In particular, the data source used to measure capital gains prior to 1885 uses averaged price data which can artificially induce serial correlation in returns (see Working (1960)). Also see the discussion of this point in Schwert (1990).

24 month averaging horizon for the large firm portfolio; the point estimate of the slope coefficient in Equation (1) is -0.74 and the t-statistic is -3.72, which has a joint significance level of 0.001. The test based on the gamma statistics provides even stronger results; here the joint significance level is 6.78×10^{-7}. Similarly, the chi-square test rejects serial independence with joint significance of 2.75×10^{-7}. These results provide direct evidence for the existence of mean reversion in large firms.

3.9 January

Jegadeesh (1991) finds strong evidence that the mean reversion present in the equally-weighted index is most pronounced in January. For the value-weighted index, the evidence in Jegadeesh (1991) of a January effect in mean reversion is weaker. The largest magnitude t-statistic concerning the predictability of January returns in the value-weighted index across all lags and samples reported in his study is -1.81.

We find additional evidence supporting the conclusion in Jegadeesh (1991) that mean reversion occurs most strongly in January. Table 9 reports the result from regressing the January excess return in year t on $\bar{r}_{t-1}^{(2)}$ for the value-weighted index in the postwar sample; the t-statistic is -3.03, and the R^2 is 16.3% ($p<.004$) which is exceptionally large given that the dependent variable is a monthly return. Table 9 also reports the results from estimating Equation (2) with the January returns omitted from the lefthand side variable. The t-statistic for the mean reversion coefficient is -1.51; this result can be contrasted with the results in Jegadeesh (1991) which finds the largest magnitiude t-statistic across all lags and samples for predicting non-January returns to be +0.88 (with a sign contrary to mean reversion).[26]

There are some methodological differences between our work and Jegadeesh (1991). The analysis in Jegadeesh uses nominal returns; our study uses excess returns. Jegadeesh reports White (1980) t-statistics. We report OLS t-statistics. To better compare our results to those in Jegadeesh (1991), we re-estimated our analysis using <u>nominal</u> returns for the value-weighted index and calculated White t-statistics. When the lefthand side variable

[26]See Table II in Jegadeesh (1991). The t-statistic, which across all lags and samples reported in Jegadeesh (1991) provides the greatest support for mean reversion, is -0.59.

contains only the return in January, the White t-statistic becomes -3.16. When the lefthand side variable is the annual return excluding January, the White t-statistic jumps to -3.12, providing strong evidence for mean reversion outside of January. While we find support for Jegadeesh's conclusion that mean reversion is strongest in January, we also find that approximately 59% of annual mean reversion occurs outside the month of January.

4. Conclusion

In the past ten years, several studies have examined serial correlation in the returns on US stock indices. Some of these studies conclude that returns exhibit negative serial correlation (mean reversion) over samples which include the Depression years and World War II (e.g. Fama and French (1988), Kim, Nelson and Startz (1991), Jegadeesh (1991)). However, these studies find little evidence of mean reversion in the postwar period. Also no evidence for mean reversion has been found in portfolios of large firms or the value-weighted index. These two characteristics of past findings suggested that mean reversion may have been a peculiarity of the Depression years and, particularly, of small firms during the Depression.

We find highly significant evidence of mean reversion in the value-weighted index over the postwar period (1947-1995). Thus mean reversion appears to be a relevant feature of the current stock market and does not appear to be restricted to small firms. We also find that the negative serial correlation occurs at a higher frequency than that found in previous studies that examined serial correlation in monthly data. We find that the average of the past 24 monthly returns is the best predictor of future returns; this differs from the findings of earlier researchers, who report the cumulative past three to five year return to be the best predictor.[27] The results are economically significant in magnitude: 48.3% of a marginal

[27] This understates the difference somewhat. We use the average of the past 24 monthly returns to predict the return in the current month. Thus the average timing of the returns in the predictor is 12.5 months away from the return being predicted. Fama and French (1988) used past cumulative returns to predict future cumulative returns. Consider the case in which the past four year return is used to predict the upcoming four year return. Here the average distance in time between the RHS and LHS returns is four years. Jegadeesh (1991) found that strongest predictability in the case in which the previous 84 monthly returns are averaged to predict the return in the current month. Here the average timing of the returns in the predictor is 48.5 months from the return being predicted, which is similar to the Fama-French findings.

shock to monthly returns will be reversed in the cumulative return over the following two years. Some of this reversal will also be reversed. But the final effect is that 39.3% of a marginal shock is reversed in the subsequent infinite horizon cumulative return.

We attribute these findings to the enhanced power of utilizing a GARCH error specification to correct for the well-documented heteroskedasticity in the monthly return series. We then use annual data to test the ability of the average of the past 24 monthly returns to predict the return in the upcoming year. Since there is no evidence of heteroskedasticity in annual returns, the estimation is conducted with OLS. Finally, a series of out-of-sample tests (US data prior to 1926 and foreign stock markets) are conducted using OLS on annual data. These results confirm the initial findings.

Appendix A. Proofs of Propositions 1 and 2

Let $Er_t = \mu$, $Var(r_t) = \sigma^2$, $m_o = \frac{1}{T}\sum_{t=1}^{T} r_t$, and $m(k) = \frac{1}{T}\sum_{j=1}^{T} [\bar{r}_{t-1}(k)]$.

m_o is the sample mean of the return series, and $m(k)$ is the sample mean of $\bar{r}_{t-1}^{(k)}$.

Lemma 1. Let $z_1 = \frac{1}{T}\sum_{t=1}^{T} [r_t - m_o][\bar{r}_{t-1}^{(k_1)} - m(k_1)]$ and $z_2 = \frac{1}{T}\sum_{t=1}^{T} [r_t - m_o][\bar{r}_{t-1}^{(k_2)} - m(k_2)]$.

Under the null that r_t is iid, $T\, Ez_1z_2 \rightarrow \frac{\sigma^4}{k_2}$ where $k_1 \leq k_2$.

Proof of Lemma 1. First we shall establish some preliminary results:

Observation 1. For $|s - t| \geq k_2 + 1$,

$E\left\{ [r_t - m_o][r_s - m_o][\bar{r}_{t-1}^{(k_1)} - m(k_1)][\bar{r}_{s-1}^{(k_2)} - m(k_2)] \right\} \rightarrow 0$ at rate T^{-4}. /

Observation 2. For $s \neq t$ and $|s-t| < k_2 + 1$,

$E\left\{ [r_t - m_o][r_s - m_o][\bar{r}_{t-1}^{(k_1)} - m(k_1)][\bar{r}_{s-1}^{(k_2)} - m(k_2)] \right\} \rightarrow 0$ at rate T^{-2}. /

Observation 3. There are only order T terms in z_1z_2 such that $|s-t| < k_2 + 1$. /

Observation 4. $E[(\bar{r}_{t-1}^{(k_1)} - m(k_1))(\bar{r}_{t-1}^{(k_2)} - m(k_2)) \mid r_t - m_o] \rightarrow \frac{\sigma^2}{k_2}$.

Proof of Observation 4.

$E[(\bar{r}_{t-1}^{(k_1)} - m(k_1))(\bar{r}_{t-1}^{(k_2)} - m(k_2)) \mid r_t - m_o] =$

$$E\left\{ (\bar{r}_{t-1}^{(k_1)} - m(k_1)) \left[\frac{k_1}{k_2}(\bar{r}_{t-1}^{(k_1)} - m(k_1)) + \frac{1}{k_2}\sum_{j=k_1+1}^{k_2} (r_{t-j} - m_o) + \frac{k_1}{k_2} m(k_1) + \frac{m_o}{k_2} - m(k_2)\right] \mid r_t - m_o \right\}.$$

Therefore,

$E[(\bar{r}_{t-1}^{(k_1)} - m(k_1))(\bar{r}_{t-1}^{(k_2)} - m(k_2)) \mid r_t - m_o] \rightarrow \frac{k_1}{k_2}\frac{\sigma^2}{k_1} + 0 + 0.$ /

The lemma can now be proven:

$$Ez_1z_2 = E\left[\frac{1}{T}\sum_{t=1}^{T} [r_t - m_o][\bar{r}_{t-1}^{(k_1)} - m(k_1)] \frac{1}{T}\sum_{s=1}^{T} [r_s - m_o][\bar{r}_{s-1}^{(k_2)} - m(k_2)] \right]$$

From Observations 1 to 3, we can ignore terms in the summation such that $t \neq s$.

$$Ez_1z_2 = \frac{1}{T^2} E\left[T [r_t - m_o]^2 [\bar{r}_{t-1}^{(k_1)} - m(k_1)][\bar{r}_{t-1}^{(k_2)} - m(k_2)]\right]$$
$$= \frac{1}{T} E\left\{ [r_t - m_o]^2 E\left[[\bar{r}_{t-1}^{(k_1)} - m(k_1)][\bar{r}_{t-1}^{(k_2)} - m(k_2)] \mid r_t - m_o\right] \right\}.$$

From Observation 4,

$T\, Ez_1z_2 \rightarrow \frac{\sigma^2}{k_2} E[r_t - m_o]^2 = \frac{\sigma^4}{k_2}$. This establishes the lemma. //

Proof of Proposition 1. Let $\hat{\beta}(k)$ be the OLS estimate of $\beta(k)$.

$$\hat{\beta}(k) = \frac{1}{T}\sum_{t=1}^{T} [r_t - m_o]\,[\bar{r}_{t-1}^{(k)} - m(k)] \;\Big/\; \frac{1}{T}\sum_{t=1}^{T} [\bar{r}_{t-1}^{(k)} - m(k)]^2$$

Consider calculating $\hat{\beta}(k)$ for $k=k_1$ and $k=k_2$. Without loss of generality, $k_1 \le k_2$. Let $\mathbf{x}$ be a 4 by 1 vector in which $x_1 = \frac{1}{T}\sum_{t=1}^{T} [r_t - m_o][\bar{r}_{t-1}^{(k_1)} - m(k_1)]$, $x_2 = \frac{1}{T}\sum_{j=1}^{T} [\bar{r}_{t-1}^{(k_1)} - m(k_1)]^2$, $x_3 = \frac{1}{T}\sum_{t=1}^{T} [r_t - m_o][\bar{r}_{t-1}^{(k_2)} - m(k_2)]$, $x_4 = \frac{1}{T}\sum_{j=1}^{T} [\bar{r}_{t-1}^{(k_2)} - m(k_2)]^2$. Consider the function $y(\mathbf{x})$: $\mathcal{R}^4 \to \mathcal{R}^2$ where $y(\mathbf{x}) = [x_1/x_2,\, x_3/x_4]$. Note that the sample estimates $[\hat{\beta}(k_1), \hat{\beta}(k_2)] = y(\mathbf{x})$. Let $E\mathbf{x} \to \boldsymbol{\phi}$ and $\mathrm{Cov}(\mathbf{x}) \to \boldsymbol{\Sigma}$. Also let $\mathbf{J} = \partial y/\partial \mathbf{x}$. The variance of $\mathbf{x}$ goes to zero at rate T^{-1}. From asymptotic theory for functions of random variables (e.g. Kendall (1980)), $\mathrm{Cov}[\hat{\beta}(k_1), \hat{\beta}(k_2)] \to \mathbf{J\Sigma J'}$ where $\mathbf{J}$ is evaluated at $\mathbf{x}=\boldsymbol{\phi}$. Note that $\boldsymbol{\phi} = [0,\, \sigma^2/k_1,\, 0,\, \sigma^2/k_2]$ under the null. Also note that

$$\mathbf{J} = \begin{bmatrix} 1/x_2 & -x_1/x_2^2 & 0 & 0 \\ 0 & 0 & 1/x_4 & -x_3/x_4^2 \end{bmatrix}, \text{ and } \mathbf{J} \text{ evaluated at } \mathbf{x}=\boldsymbol{\phi} \text{ is } \begin{bmatrix} k_1/\sigma^2 & 0 & 0 & 0 \\ 0 & 0 & k_2/\sigma^2 & 0 \end{bmatrix}.$$

$$\mathbf{J\Sigma J'} = \frac{1}{\sigma^4}\begin{bmatrix} k_1^2\Sigma_{11} & k_1k_2\Sigma_{13} \\ k_1k_2\Sigma_{31} & k_2^2\Sigma_{33} \end{bmatrix}.$$ From Lemma 1: $\mathrm{T}\Sigma_{11} = \sigma^4/k_1$, $\mathrm{T}\Sigma_{13} = \mathrm{T}\Sigma_{31} = \sigma^4/k_2$, and $\mathrm{T}\Sigma_{33} = \sigma^4/k_2$. This establishes the first 2 results in the proposition: $T\,\mathrm{Var}(\hat{\beta}(k)) = k$ and $\rho(\hat{\beta}(k_1), \hat{\beta}(k_2)) = \dfrac{k_1}{\sqrt{k_1\, k_2}}$. Since the asymptotic standard deviation of $\hat{\beta}(k)$ is $\sqrt{k/T}$, $t(k)$ converges to $\hat{\beta}(k)\sqrt{T/k}$ asymptotically. The remaining 2 results concerning the asymptotic joint distribution of the OLS t-statistics immediately follow. //

Proof of Proposition 2. Let $\hat{\boldsymbol{\beta}}$ be the N by 1 vector $(\hat{\beta}(k_1), \ldots, \hat{\beta}(k_N))$. $\mathrm{Cov}(\mathbf{c}'\hat{\boldsymbol{\beta}}, \mathbf{d}'\hat{\boldsymbol{\beta}}) \to \mathbf{c}'\boldsymbol{\Sigma}\mathbf{d}$ where $\boldsymbol{\Sigma}$ is the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$. For $n > 1$, $\hat{\gamma}(n) = (\mathbf{e}_n - \mathbf{e}_{n-1})'\hat{\boldsymbol{\beta}}$ where $\mathbf{e}_n$ is the nth standard basis vector. For $n = 1$, $\hat{\gamma}(n) = \mathbf{e}_n'\hat{\boldsymbol{\beta}}$. Without loss of generality, assume $m \le n$. From Proposition 1, $\Sigma(n,n) = n\kappa/T$ and $\Sigma(n,m) = m\kappa/T$. We proceed by considering 4 cases.

Case 1: $m = n$, $m > 1$. Here $\mathbf{c}'\boldsymbol{\Sigma}\mathbf{d} = \Sigma(n,n) - \Sigma(n-1,n) - [\Sigma(n,n-1) - \Sigma(n-1,n-1)] = n\kappa/T - (n-1)\kappa/T - (n-1)\kappa/T + (n-1)\kappa/T = \kappa/T$.

Case 2: $m = n$, $m = 1$. Here $\mathbf{c}'\mathbf{\Sigma d} = \Sigma(1,1) = \kappa/T$.

Therefore, $T\,\mathrm{Var}(\hat{\gamma}_{(n)}) = \kappa$ for all n.

Case 3: $m \leq n+1$, $m > 1$. Here $\mathbf{c}'\mathbf{\Sigma d} = \Sigma(m,n) - \Sigma(m-1,n) - \left[\Sigma(m,n-1) - \Sigma(m-1,n-1)\right] =$

$m\kappa/T - (m-1)\kappa/T - m\kappa/T + (m-1)\kappa/T = 0$.

Case 4: $n \geq 2$, $m = 1$. Here $\mathbf{c}'\mathbf{\Sigma d} = \Sigma(1,n) - \Sigma(1,n-1) = (1)\kappa/T - (1)\kappa/T = 0$.

Therefore, $\mathrm{Cov}(\hat{\gamma}_{(m)}, \hat{\gamma}_{(n)}) = 0$ for $m \neq n$. //

Appendix B. Percentage Cumulative Reversal

This appendix derives the formula used to in Section 3.3 to calculate the percentage of a one-time marginal shock which is reversed in the subsequent infinite horizon cumulative return. Consider the model $r_t - r_f = \beta_0 + \beta \frac{1}{K} \sum_{j=1}^{k} r_{t-j} + e_t$ where r_f is the risk-free rate (assumed to be constant), and consider a one-time shock of size u that occurs at time 0. The model implies that the unconditional expectation of r is $(r_f+\beta_0)/(1-\beta)$. Therefore, a one-time shock u that occurs at time 0 implies that $r_t = (r_f+\beta_0)/(1-\beta)$ for $t < 0$ and $r_0 = u + (r_f+\beta_0)/(1-\beta)$.

Let R(u,T) be percentage of the time zero shock which is reversed in the cumulative return from time 1 to time T. Along the path with no time 0 shock, the cumulative return from time 1 to time T is $(1 + \frac{r_f+\beta_0}{1-\beta})^T$. Therefore, the part of the cumulative return from time 1 to time T that is attributable to the time 0 shock is the difference $\exp\{\sum_{t=1}^{T} \ln(1+r_t)\} - (1 + \frac{r_f+\beta_0}{1-\beta})^T$. However, this value occurs at time T, while the value of the initial shock u ocurs at time 1. Therefore, R(u,T) equals

$$\frac{\exp\{\sum_{t=1}^{T} \ln(1+r_t)\} - (1 + \frac{r_f+\beta_0}{1-\beta})^T}{u\,(1 + \frac{r_f+\beta_0}{1-\beta})^T}.$$

Let $q_t = r_t - (r_f+\beta_0)/(1-\beta)$. The model can be expressed as $q_t = \beta \frac{1}{K} \sum_{j=1}^{k} q_{t-j} + e_t$ and the one-time shock implies $q_t = 0$ for $t < 0$ and $q_t = u$ for $t = 0$. Note that $\sum_{t=1}^{\infty} q_t - \sum_{t=-k+1}^{\infty} \beta q_t = 0$. Also $\sum_{t=-\infty}^{0} q_t - \sum_{t=-\infty}^{-k} \beta q_t = q_0$. Hence, $\sum_{t=-\infty}^{\infty} q_t - \beta \sum_{t=-\infty}^{\infty} q_t = q_0$. Let $S = \sum_{t=-\infty}^{\infty} q_t$. Since $q_0 = u$, we have the equation $(1-\beta)S = u$ which has the solution $S = u/(1-\beta)$. Let $Q(u,T) = \sum_{t=1}^{T} q_t$. Then, $\lim_{T\to\infty} Q(u,T) = S - q_0 = u/(1-\beta) - u = u\beta/(1-\beta)$.

We are interested in $\lim_{\substack{u\to 0\\ T\to\infty}} R(u,T)$ and can evaluate the limits in either order. By l'Hopital's rule, $\lim_{u\to 0} R(u,T) = \frac{1-\beta}{1 + r_f + \beta_0 - \beta} \frac{dQ(u,T)}{du}$. Since Q(u,T) is continuously differentiable in u and $\lim_{T\to\infty} Q(u,T) = u\beta/(1-\beta)$, $\lim_{T\to\infty} \lim_{u\to 0} R(u,T) = \frac{\beta}{1 + r_f + \beta_0 - \beta}$.

References

Attanasio, Orazio P., 1991, Risk, Time Varying Second Moments and Market Efficiency, Review of Economic Studies 58, 479-494.

Bassett, G. and R. Koenker, 1978, Asymptotic Theory of Least Absolute Error Regressions, Journal of the American Statistical Association 73, 618-622.

Bodurtha, James N. and Nelson C. Mark, 1991, Testing the CAPM with Time Varying Risks and Returns, Journal of Finance 46, 1485-1505.

Bernard, Victor L. and Jacob K. Thomas, 1990, Evidence That Stock Prices Do Not Fully Reflect The Implications of Current Earnings for Future Earnings, Journal of Accounting and Economics 13, 305-340.

Bidarkota, Prasad and J. Huston McCulloch, 1996, Real Stock Returns: Non-normality, Seasonality, and Volatility Persistence, But No Predictability, Working paper, Ohio State University.

Bollerslev, Tim, Ray Chou, and Kenneth Kroner, 1992, ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence, Journal of Econometrics 52, 5-59.

Daniel, Kent, 1994, The Power and Size of Mean Reversion Tests, Working paper, Northwestern University.

Daniel, Kent and Walter Torous, 1995, Common Stock Returns and the Business Cycle, Working paper, Northwestern University.

Fama, Eugene F. and Kenneth R. French, 1988, Permanent and Temporary Components of Stock Prices, Journal of Political Economy 96, 246-273.

French, Kenneth R., G. William Schwert, and Robert F. Stambaugh, 1987, Expected Stock Returns and Volatility, Journal of Financial Economics 19, 3-29.

Hansen, Lars P., 1982, Large Sample Properties of Generalized Method of Moments Estimators, Econometrica 50, 1029-1054.

Hodrick, Robert, 1992, Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, Review of Financial Studies 5, 357-386.

Jegadeesh, Narasimhan, 1991, Seasonality in Stock Price Mean Reversion: Evidence from the U.S. and the U.K., Journal of Finance 46, 1427-1444.

Kendall, M. and A. Stuart, 1977, The Advanced Theory of Statistics, 4th edition, Macmillan Publishing Co., NY

Kim, Myung J., Charles R. Nelson, and Richard Startz, 1991, Mean Reversion in Stock Prices? A Reappraisal of the Empirical Evidence, Review of Economic Studies 58, 515-528.

Lamoureux, Christopher G. and Guofu Zhou, 1996, Temporary Components of Stock Returns: What Do the Data Tell Us?, Review of Financial Studies 9, 1033-1059.

McCusk, John J., 1992, "How Much Is That in Real Money?, American Antiquarian Society.

Merton, Robert C., 1980, On Estimating the Expected Return on the Market: An Exploratory Investigation, Journal of Financial Economics 8, 323-361.

Poterba, James M. and Lawrence H. Summers, 1988, Mean Reversion in Stock Prices: Evidence and Implications, Journal of Financial Economics 22, 27-59.

Richardson, Matthew, 1993, Temporary Components of Stock Prices: A Skeptic's View, Journal of Business & Economic Statistics 11, 199-207.

Richardson, Matthew and Tom Smith, 1991, Tests of Financial Models in the Presence of Overlapping Observations, Review of Financial Studies 4, 227-254.

Richardson, Matthew and James Stock, 1989, Drawing Inferences from Statistics Based on Multiyear Asset Returns, Journal of Financial Economics 25, 323-348.

Schwert, William G., 1990, Indexes of U.S. Stock Prices from 1802 to 1987, Journal of Business 63, 399-426.

Summer, Lawrence H., 1986, Does the Stock Market Rationally Reflect Fundamental Values?, Journal of Finance 41, 591-601.

White, Hal, 1980, A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, Econometrica 48, 817-838.

Working, Holbrook, 1960, A Note on the Correlation of First Differences of Averages in Random Chains, Econometrica 28, 916-918.

Table 1

Estimation on monthly data from January 1947 to December 1995 of

$$r_t = \beta_0 + \beta\, \bar{r}^{(k)}_{t-1} + e_t$$

$$Var(e_t) = h_t$$

$$h_t = \alpha_0 + \alpha_1 h_{t-1} + \alpha_2 h_{t-2} + \alpha_3 h_{t-3} + \phi_1 e^2_{t-1} + \phi_2 e^2_{t-2} + \phi_3 e^2_{t-3}$$

$$\alpha_1 + \alpha_2 + \alpha_3 + \phi_1 + \phi_2 + \phi_3 = 1 \quad \text{(IGARCH constraint)}$$

$$k \in \{12, 24, 36, 48, 60, 72, 84\}$$

r_t is the excess return on the value-weighted index in month t.

k is the number of lagged excess returns used to form predictor in mean equation.

Reported value of k is the value that maximizes the likelihood function.

All significance levels are two-sided.

Mean Equation

Variable	β_0	β	k
Estimate	.011	-.66	24
T-Statistic	5.11	-3.26*	
P-Value (univariate)	.00	.0011**	

Variance Equation

Variable	α_0	α_1	α_2	α_3	ϕ_1	ϕ_2	ϕ_3
Estimate	8.01 x10^{-5}	.60	.10	.03	.074	-.018	.22
T-Statistic	2.19		.22	.10	1.61	-.24	2.74
P-Value	.03		.82	.92	.11	.81	.006

*This is a theoretical t-ratio formed by dividing the point estimate by the asymptotic standard deviation of the estimate under the null as derived in Proposition 1. The empirical t-ratio reported from the maximum likelihood procedure is -3.30; the two numbers differ by only four one-hundredths.

**This p-value corresponds to the theoretical t-ratio, -3.26; the p-value for the empirical t-ratio, -3.30, is .00098.

Table 2

Correlation of $\hat{\beta}_{(k)}$ Estimates Under the Null

	$\hat{\beta}_{(12)}$	$\hat{\beta}_{(24)}$	$\hat{\beta}_{(36)}$	$\hat{\beta}_{(48)}$	$\hat{\beta}_{(60)}$	$\hat{\beta}_{(72)}$	$\hat{\beta}_{(84)}$
$\hat{\beta}_{(12)}$	1	.707	.577	.5	.447	.408	.378
$\hat{\beta}_{(24)}$		1	.816	.707	.632	.577	.535
$\hat{\beta}_{(36)}$			1	.866	.775	.707	.655
$\hat{\beta}_{(48)}$				1	.894	.816	.756
$\hat{\beta}_{(60)}$					1	.913	.845
$\hat{\beta}_{(72)}$						1	.926
$\hat{\beta}_{(84)}$							1

This table calculates the correlations under the null hypothesis that monthly excess returns are serially uncorrelated.

Table 3

Results for the Seven Independent Test Statistics $\hat{\gamma}(12)$ to $\hat{\gamma}(84)$

Point estimates and t-statistics for $\hat{\gamma}(k)$ where $\hat{\gamma}(k)$ is defined as follows.

$$\hat{\gamma}(k_n) = \hat{\beta}(k_n) - \hat{\beta}(k_{n-1}) \qquad \text{for } 1 < n \leq N$$
$$= \hat{\beta}(k_n) \qquad \text{for } n = 1.$$

The point estimates for $\hat{\beta}(k)$ come from maximum likelihood estimation of Equation (1) where k is treated as an exogenous model parameter. The standard errors used to form the t-statistics are derived in Proposition 2.

k	$\hat{\beta}$	$\hat{\gamma}$	t-Statistic on $\hat{\gamma}$
12	-0.01	-0.01	-0.08
24	-0.66	-0.65	-4.52
36	-0.60	0.06	0.40
48	-0.42	0.18	1.29
60	-0.02	0.39	2.75
72	-0.01	0.01	0.09
84	0.21	0.22	1.55

Table 4
Theoretical Versus Empirical T-Statistics and Standard Errors

The model

$$r_t = \beta_0 + \beta\,\bar{r}^{(k)}_{t-1} + e_t$$

$$Var(e_t) = h_t$$

$$h_t = \alpha_0 + \alpha_1 h_{t-1} + \alpha_2 h_{t-2} + \alpha_3 h_{t-3} + \phi_1 e^2_{t-1} + \phi_2 e^2_{t-2} + \phi_3 e^2_{t-3}$$

$$\alpha_1 + \alpha_2 + \alpha_3 + \phi_1 + \phi_2 + \phi_3 = 1 \quad \text{(IGARCH constraint)}$$

was estimated for values of k in the set {12, 24, 36, 48, 60, 72, 84} on monthly data from January 1947 to December 1995 where
r_t is the excess return on the value-weighted index in month t.
k is the number of lagged excess returns used to form predictor in mean equation.

$\hat{\beta}$ and "Empirical Standard Error" are the maximum likelihood estimate and the maximum likelihood standard error. "Theoretical Standard Error" is calculated from Proposition 1. The "Theoretical T-Statistic" is the point estimate divided by the theoretical standard error. The "Empirical T-Statistic" is the point estimate divided by the empirical standard error.

k	$\hat{\beta}$	Empirical SE	Theoretical SE	Empirical T-Statistic	Theoretical T-Statistic
12	-0.01	.13	.14	-0.09	-0.08
24	-0.66	.20	.20	-3.30	-3.26
36	-0.60	.28	.25	-2.17	-2.43
48	-0.42	.32	.29	-1.30	-1.46
60	-0.02	.32	.32	-0.07	-0.07
72	-0.01	.31	.35	-0.03	-0.03
84	0.21	.33	.38	0.64	0.56

Table 5

Response to A One-Time Shock

This table demonstrates the effect of an infinitesimal one-time shock at time 0 to the subsequent path of the expected returns process in the model $r_t - r_f = \beta_o + \beta\,\bar{r}^{(24)}_{t-1} + e_t$ where β_o and β are set equal to the values estimated in Table 1 and the risk-free rate is assumed to be constant and set equal to its sample average. Along the path in which there is no time 0 shock, the return is $(r_f+\beta_o)/(1-\beta)$ each month. Let $q_t = r_t - (r_f+\beta_o)/(1-\beta)$. For $t > 0$, q_t is the shift that the time 0 shock induces to the return in month t.

Column 2 in the table reports $\lim_{u\to 0} q_t/u$ where u is the time 0 shock.

Column 3 reports the sum of the previous 24 values of $\lim_{u\to 0} q_t/u$ which is analogous to $\bar{r}^{(24)}_{t-1}$ for an infinitesimal shock.

The cumulative reversal measures the percentage of a time zero infinitesimal shock which has been reversed by the cumulative return from time 1. The percentage of a time zero shock u which has been reversed by month T is

$$R(u,T) = \frac{\exp\{\sum_{t=1}^{T} \ln(1+r_t)\} - (1+\frac{r_f+\beta_o}{1-\beta})^T}{u\,(1+\frac{r_f+\beta_o}{1-\beta})^T}$$

where $(1+\frac{r_f+\beta_o}{1-\beta})^T$ is the cumulative return from time 1 to time T in the case in which there was no time 0 shock.

Column 4 reports $\lim_{u\to 0} R(u,T)$. Time is measured in months.

t	Shift to r_t (percent of initial shock)	Shift to $\bar{r}^{(24)}_{t-1}$ (percent of initial shock)	Cumulative Reversal (percent of initial shock)
1	-2.7	4.2	2.7
2	-2.7	4.1	5.4
3	-2.6	3.9	7.9
6	-2.4	3.6	15.2
9	-2.2	3.3	21.9
12	-2.0	3.1	28.1
15	-1.9	2.8	33.8
18	-1.7	2.6	39.0
21	-1.6	2.4	43.8
24	-1.4	2.2	48.3
27	1.1	-1.7	44.6
30	0.8	-1.3	41.9
33	0.6	-0.9	39.9
36	0.4	-0.5	38.7
42	0.01	-0.02	37.8
48	-0.2	0.4	38.6
54	-0.1	0.1	39.6
60	0.02	-0.03	39.7
72	0.03	-0.04	39.3
84	-0.01	0.02	39.3

Table 6

Annual Results

Panel A:
Estimation using OLS on **annual** data from 1947 to 1995 of

$$r_t = \beta_0 + \beta \bar{r}_{t-1}^{(2)} + e_t.$$

r_t is the excess return on the value-weighted index in year t.
Annual excess returns are formed by compounding monthly excess returns.

Variable	β_0	β
Estimate	0.12	-0.46
T-Statistic	4.23	-2.21
P-Value (2-Sided)	.000	.03
$R^2 = .094$		

Panel B:
Estimation using OLS on **annual** data from 1947 to 1995 of

$$r_t = \beta_0 + \beta_1 \bar{r}_{t-1}^{(2)} + \beta_2 I7475_t + e_t.$$

$I7475_t$ is a dummy variable indicating whether year t is in the period 1974-75.
r_t is the excess return on the value-weighted index in year t.
Annual excess returns are formed by compounding monthly excess returns.

Variable	β_0	β_1	β_2
Estimate	0.15	-0.70	-0.27
T-Statistic	4.92	-3.06	2.20
P-Value (2-Sided)	.000	.004	.03
$R^2 = .180$			

Panel C:
Estimation using **Least Absolute Deviations** on annual data from 1947 to 1995 of

$$r_t = \beta_0 + \beta \bar{r}_{t-1}^{(2)} + e_t.$$

r_t is the excess return on the value-weighted index in year t.
Annual excess returns are formed by compounding monthly excess returns.

Variable	β_0	β
Estimate	0.13	-0.56
T-Statistic	3.88	-2.18
P-Value (2-Sided)	.000	.035

Table 7

Out-Of-Sample Analysis: Foreign Stock Markets. US Market Prior to 1925

Panel A:

Estimation using OLS on **annual** data from 1951 to 1995 of

$$r_t = \beta_0 + \beta\,\bar{r}^{(2)}_{t-1} + e_t$$

r_t is the the real return as reported in the DRI Basic Economics database in year t.
The real return for year t is percentage change in the real stock price in year t.

Country	β	T-Statistic	Significance (2-Sided)
Canada	-0.39	-1.87	.06
France	-0.17	-0.73	.47
Japan	-0.23	-1.11	.27
UK	-0.48	-2.13	.04

Panel B:

Estimation using OLS on annual US data from **1885 to 1925** of

$$r_t = \beta_0 + \beta\,\bar{r}^{(2)}_{t-1} + e_t$$

r_t is the real return on the index described in Schwert (1990) in year t.
The real return for year t is computed as the nominal return for year t minus the percentage change in the price level in year t.

Variable	β_0	β
Estimate	0.16	-0.54
T-Statistic	3.47	-2.38
P-Value (2-Sided)	.001	.02

Table 8
Large Firms

This table replicates many of the items reported in Tables 1, 3, and 6 for an equally-weighted portfolio consisting of the largest quintile of firms by market value of equity. Each of the procedures below is estimated from January 1947 to December 1995. Procedures 1 through 3 are estimated on monthly returns.

Procedure 1: All parameters of Equation (1) are estimated jointly by maximum likelihood. The reported β corresponds to the value of k that maximizes the model's likelihood function. The standard errors used to form the t-statistics are derived in Proposition 1.

Procedure 2: $\gamma(k)$ is calculated for seven averaging horizons ranging from 12 months to 84 months. The value of γ with the maximal t-statistic is reported below. The standard errors used to form the t-statistics are derived in Proposition 2.

Procedure 3: For each of seven averaging horizons ranging from 12 months to 84 months, k was held constant and the remaining parameters of Equation (1) were estimated by maximum likelihood. A chi-square test was performed on the seven resulting estimated slope coefficients for the covariance matrix derived in Proposition 1.

Procedure 4: The equation $r_t = \beta_0 + \beta \bar{r}^{(2)}_{t-1} + e_t$ is estimated using OLS on annual data.

	Portfolio	Largest Size Quintile	Value-Weighted Index
k	estimate	24	24
GARCH β (monthly data)	estimate	-0.74	-0.66
	t-statistic	-3.72	-3.26
	joint significance (2-sided)	.001	.007
γ	estimate	-0.71	-0.67
	t-statistic	-4.98	-4.52
	joint significance (2-sided)	6.78×10^{-7}	4.33×10^{-5}
χ^2	test statistic	43.4	32.3
	joint significance	2.75×10^{-7}	3.57×10^{-5}
OLS β (annual data)	estimate	-0.48	-0.46
	t-statistic	-2.36	-2.21
	R^2	.106	.094
	significance (2-sided)	.02	.03

Table 9
January

Estimation using OLS on annual data from 1947 to 1995 of

$$r_t = \beta_0 + \beta \bar{r}^{(2)}_{t-1} + e_t$$

for 2 samples: one in which the lefthand side variable includes only the return in the month of January and a second in which the lefthand side variable excludes the January return. All data are excess returns on the CRSP value-weighted index.

LHS	β	T-Statistic	Significance (2-Sided)
January Only	-0.19	-3.03	.004
Feb-Dec.	-0.27	-1.51	.14

Figure 1

Mean Reversion, Annual U.S. Returns

Annual excess returns, CRSP value-weighted index, 1947-1995. Y-axis is annual excess return in year t. X-axis is average annual excess return in years t-1 and t-2. Annual excess returns are formed by compounding monthly excess returns.

Solid line is the OLS fitted regression line.

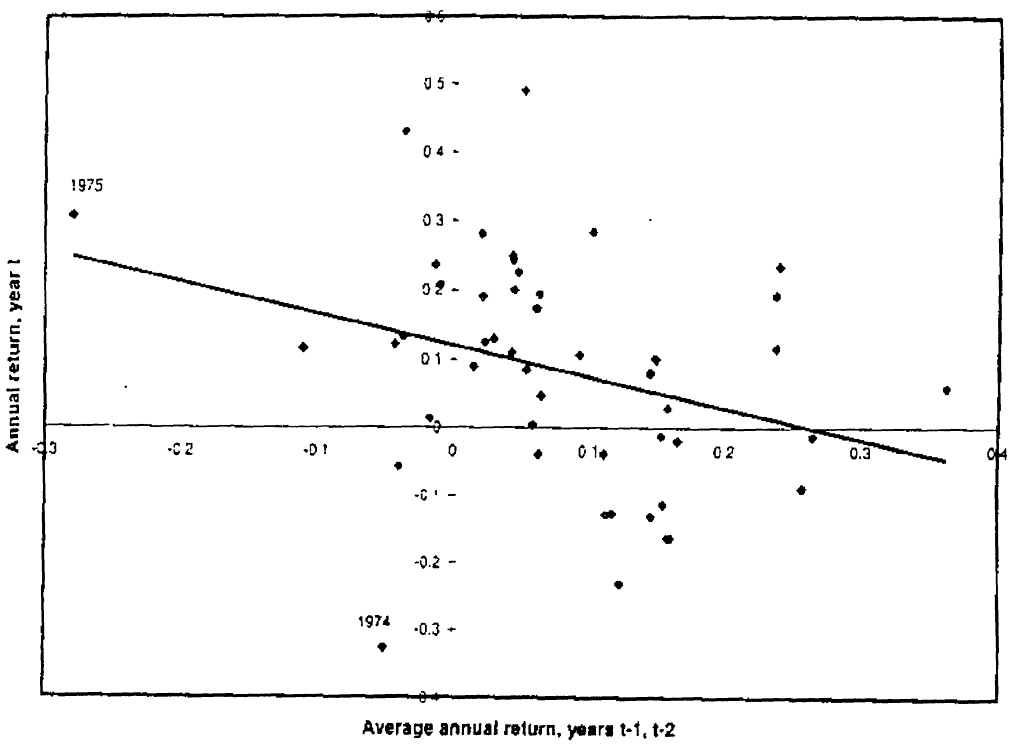

Figure 2

Mean Reversion, Annual U.K. Returns

Annual real returns. U.K. stock index, 1951-1995. Y-axis is annual real return in year t. X-axis is average annual real return in years t-1 and t-2. Annual real returns are formed by compounding monthly real returns.

Solid line is the OLS fitted regression line.

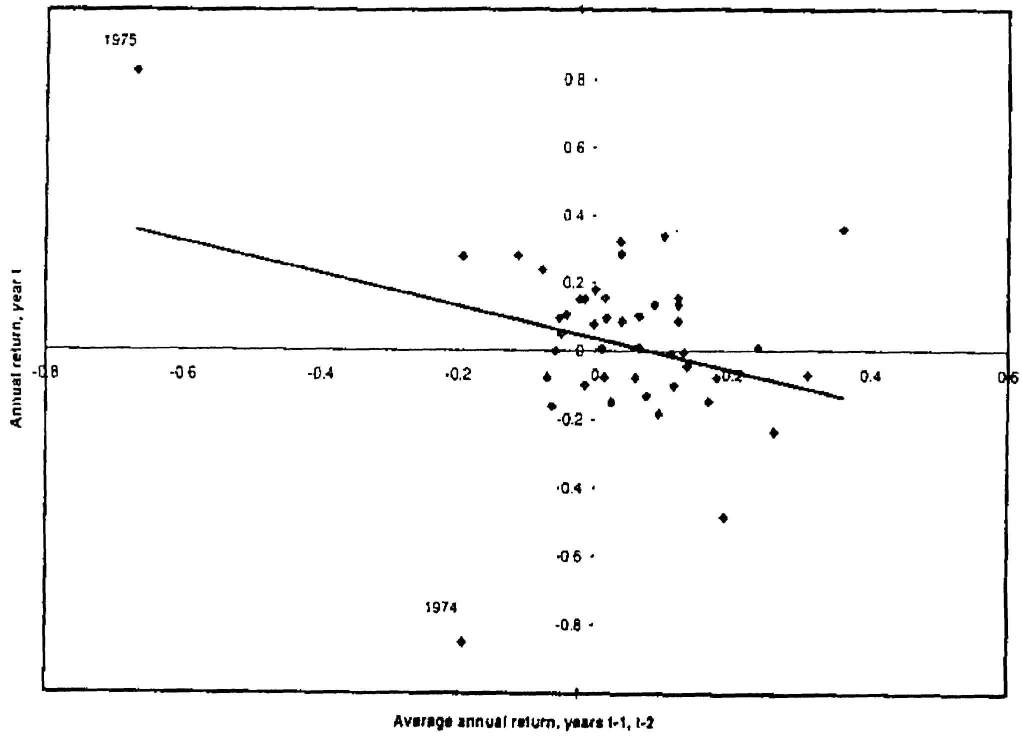

Figure 3

Response to a One-Time Unit Shock

Effect of a time 0 shock on subsequent monthly and cumulative returns, based on maximum likelihood estimates of Equation (1), value-weighted index, 1947-1995, and a monthly risk-free rate of .39%, the average value during 1947-1995.

The dotted line displays the subsequent monthly reversals of the initial shock. The solid line displays the cumulative reversal. Both are expressed as percentages of the initial shock. The asymptote for the cumulative reversal is -39.3%.