The Behavior of Random Variables with Nonstationary Variance and the Distribution of Security Prices

By

BARR ROSENBERG

Research Program in Finance Working Papers are preliminary in nature; their purpose is to stimulate discussion and comment. Therefore, they should not be cited or quoted in any publication without the permission of the author. Single copies of a paper may be requested from the Institute of Business and Economic Research.
RESEARCH PROGRAM IN FINANCE AT THE
WALTER A. HAAS SCHOOL OF BUSINESS,
UNIVERSITY OF CALIFORNIA, BERKELEY

The Research Program in Finance in the Walter A. Haas School of Business at the University of California has as its purpose the conduct and encouragement of research in finance, investments, banking, securities markets, and financial institutions. The present reprint and working paper series were established in 1971 in conjunction with a grant from the Dean Witter Foundation.

INSTITUTE OF BUSINESS AND ECONOMIC RESEARCH
Carl Shapiro, Director

The Institute of Business and Economic Research is an organized research unit at the University of California, Berkeley, whose mission is to promote research by faculty and graduate students in the fields of business and economics. The Institute carries out its mission by organizing programs and activities that enrich the research environment, administering extramural research awards, publishing working papers, and making direct grants for research.
RESEARCH PROGRAM IN FINANCE

WORKING PAPER NO. 11

THE BEHAVIOR OF RANDOM VARIABLES WITH NONSTATIONARY VARIANCE
AND THE DISTRIBUTION OF SECURITY PRICES

by

Barr Rosenberg

December 1972

Barr Rosenberg is Assistant Professor, Graduate School of Business Administration, University of California, Berkeley. This research was supported by NSF Grant GS-3306. The assistance of Michel Houglet made this experiment possible and is gratefully acknowledged.
THE BEHAVIOR OF RANDOM VARIABLES WITH NONSTATIONARY VARIANCE
AND THE DISTRIBUTION OF SECURITY PRICES

Barr Rosenberg

When the variance of a population of random variables is nonstationary, the population kurtosis is greater than the kurtosis of the probability distribution of the individual random variables. Therefore, the high kurtosis observed in the distribution of security prices can be explained by high kurtosis in the individual price changes, nonstationarity in the variances of the price changes, or any combination of these two causes. For a 100-year series of monthly changes in the Standard & Poor's Composite Index, fluctuations in variance as forecasted by a two-parameter model explain 70 percent of the deviation of the sample kurtosis from normality. The results suggest that increments in the logarithms of security prices obey a normal distribution with predictably fluctuating variance.

1. INTRODUCTION

In hundreds of economic series the empirical frequency of outliers observations is greater than would be expected from the normal distribution. For example, in Figure A, the frequency distribution of 100 years of monthly changes in the logarithm of the Standard & Poor's Composite Price Index for the New York Stock Exchange is superimposed on a graph of the normal distribution with equal variance. The empirical distribution shows a greater concentration of mass near zero and much more massive tails than the normal distribution. The largest price change is 9.3 standard deviations from the mean. With kurtosis defined as the ratio of the fourth moment to the square
of the second moment, the sample kurtosis is 14.79, in contrast to a kurtosis of 3 for the normal distribution. Unimodal leptokurtic distributions similar to this one have been observed in virtually every security or commodity market that has been studied.

As a result, research in finance has been concerned with explaining these empirical distributions and with analyzing their consequences for portfolio management. The first noteworthy innovation in this direction was the suggestion of the class of nonnormal stable distributions by Mandelbrot [11]. This hypothesis implies that the variance of security price changes is infinite and, hence, has important implications for statistical methodology (summarized recently in [8]) and for portfolio management (summarized recently in [6, p. 261-275]). Press [15] later suggested that each price change is the sum of a random number, \( r \), of normally distributed variables, where \( r \) obeys the Poisson Distribution. In this model, every price change obeys the same compound distribution equal to a normal distribution with Poisson distributed variance. Praetz [14] maintained Press's basic formulation but proposed that the variance might follow an inverted gamma distribution, and confirmed that the resulting distribution provided a better fit to some empirical data than any of the previously suggested alternatives.

All of these approaches to the problem coincide in assuming that price changes are identically distributed in all periods, in which case this distribution must be similar to the empirical frequency distribution observed over a large number of periods. Thus, all approaches preserve the basic implication of the very long tails in the empirical frequency distributions: in each and every period the investor is faced with a nonnegligible probability of a price change of extraordinary magnitude. This view has
been widely accepted by students of finance in the past decade. It has
sobering consequences. For instance, the average magnitude of the largest
1 percent of the monthly changes in the S&P Composite Index is nine times as
great as the average magnitude of the remaining 99 percent. Apparently the
investor will first be lulled by a long series of small changes in security
prices drawn from the mass of the probability distribution near zero, and
then will be suddenly shocked by an immense change drawn from the tails of the
distribution.

Formally, let $z_t$, $t=1, T$ denote a series of price changes. Following
the arguments in [1, p. 208], [12], and [13], $z_t$ should be a measure
of the proportional change in price. The choice of the particular propor-
tional measure is of minor importance for the purposes of this study, and
the traditional choice of $z_t = \log p_{t+1} - \log p_t$ will be made. Then the
prevailing view in finance is represented by the null hypothesis:

The variables $z_t$, $t=1, T$ are serially independent and
identically distributed.

The alternative hypothesis offered in this paper is that the variance
of $z_t$ fluctuates in a predictable fashion over time. This hypothesis im-
plies that the investor will not be surprised by extraordinarily large price
changes when they occur, but rather will have been forewarned by signs that
the forthcoming variance is high. Formally, the alternative hypothesis is
the following:

The random variables $z_t$ are determined by the model
$z_t = m_t + \eta_t$, $t=1, T$. The $\eta_t$ are serially inde-
pendent random variables with identical distribution
functions $F(\cdot)$ having mean equal to zero, variance
equal to one, and kurtosis equal to $\gamma$. The variables
$\eta_t$, $t=1, T$, which are the variances of the price changes,
обей a stochastic process that can be forecasted. The $\eta_t$
are contemporaneously independent of the $\eta_t$. 
A histogram of the 100-year history of monthly price changes, corrected for forecasted variance fluctuations by the methods developed below, appears in Figure B. Notice the reductions in the variance and the kurtosis of the distribution relative to Figure A. The results suggest that the distribution of the $\eta_t$ may eventually be shown to be indistinguishable from the normal distribution.

In Section 2, below, the relationship between fluctuations in the variance of a series of random variables and the kurtosis of the series is explored. In Section 3, the use of a linear regression model to analyze fluctuations in variance is explained. In Section 4, the stock price series to be analyzed is described and some necessary adjustments to the series are derived. In Section 5, the results of the experiment are reported. Section 6 provides a brief discussion of the implications of the experiment.

2. THE KURTOSIS OF A POPULATION WITH NONCONSTANT VARIANCE

2.1. Consider the case defined in the alternative hypothesis. Let the random variables $y_t = z_t - m_t$, $t = 1, \ldots, T$ be transformed to have mean zero. In studying the properties of the $y_t$, the first step is to describe their probability distribution on the assumption that the $v_t$ are equal to a set of fixed realized values. Let $E_v(\cdot) \equiv E(\cdot | v_1, \ldots, v_T)$ be the expectation operator which implements this perspective. Then it is an immediate consequence of the independence between $v_t$ and $\eta_t$, and of the assumption that $\eta_t$ has mean zero, variance 1, and kurtosis equal to $\gamma$, that

$$
\begin{align*}
E_v(y_t) &= \sqrt{v_t} E_v(\eta_t) = 0 \\
E_v(y_t^2) &= v_t E_v(\eta_t^2) = v_t \\
E_v(y_t^4) &= v_t^2 E_v(\eta_t^4) = \gamma v_t^2
\end{align*}
$$

(2.1)
\[ \text{KURTOSIS}_v(y_t) = E_v\left(\frac{y_t^4}{(E_v(y_t^2))^2}\right) = \gamma. \] (2.2)

Similarly, the expected values of the sample moments of the series \( \{y_1, \ldots, y_T\} \) are

\[
\begin{aligned}
E_v(\mu_1) &
= E_v\left(\frac{\Sigma y_t}{T}\right) = \Sigma \left(E_v(y_t) / T\right) = 0 \\
E_v(\mu_2) &
= E_v\left(\frac{\Sigma y_t^2}{T}\right) = \Sigma \left(E_v(y_t^2) / T\right) = \Sigma v_t / T = v_1 \\
E_v(\mu_4) &
= E_v\left(\frac{\Sigma y_t^4}{T}\right) = \Sigma \left(E_v(y_t^4) / T\right) = \gamma \Sigma v_t^2 / T = \gamma v_2
\end{aligned}
\] (2.3)

\[ \text{KURTOSIS}_v(y) = E_v\left(\frac{\mu_4}{(E_v(\mu_2))^2}\right) = \frac{\gamma v_2}{v_1^2}, \] (2.4)

where \( \mu_i \) is the \( i \)th moment about zero of the \( y \) sample, and \( v_i \) is the corresponding moment of the \( v \) sample. Let \( v'_1 \) be the \( i \)th moment of the \( v \) sample about the mean, so that the equality \( v_2 = v'_2 + v_1^2 \) states the familiar fact that the mean square is equal to the variance plus the square of the mean. Furthermore, let \( c = \sqrt{v'_2 / v_1^2} \) be the coefficient of variation of the \( v \) sample. Then,

\[ \text{KURTOSIS}_v(y) = ((v'_2 + v_1^2) / v_1^2)\gamma = (1 + c^2)\gamma. \] (2.5)

Thus, the population kurtosis of \( y \) is always greater than the kurtosis of the individual distributions from which the population was compounded, and the increase is determined by the coefficient of variation of \( v \).

Also, under appropriate assumptions concerning the existence of moments for \( F \) and the moments of the \( v \) sample, it can be shown that
A. HISTOGRAM OF STANDARDIZED MONTHLY PRICE CHANGES$^d$ COMPARED TO A STANDARD NORMAL DISTRIBUTION

![Histogram of standardized monthly price changes compared to a standard normal distribution.](image)

a. The series $y_t$ as defined in Section 4, divided by $\sqrt{\frac{\sum_{t=1}^{T} y_t^2}{T}}$.

B. HISTOGRAM OF MONTHLY PRICE CHANGES CORRECTED FOR FORECAST STANDARD ERRORS$^a$

![Histogram of monthly price changes corrected for forecast standard errors.](image)

a. The series $(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{y_t^2}{T})/\hat{s}_t$, where $\hat{s}_t$ is the (standardized) forecast standard deviation, explained in Section 5. The histogram is superimposed on the same standard normal distribution as in Figure A.
the probability limit (plim) of the sample kurtosis \( \mu_4/\mu_2^2 \) as \( T \to \infty \) is equal to the population kurtosis.

2.2. The next step is to examine the probability distribution of \( y \) on the assumption that the variances \( \{v_1, \ldots, v_T\} \) are the realization of a stochastic process \( V \). Let \( \bar{\nu} = E[v_1] \) and \( \sigma_v^2 = E[v_1^2] \) denote the (finite) expected values of the sample mean and sample variance of the realizations of \( V \), and let the pattern of serial dependence in \( V \) remain unspecified. (As a special case, \( V \) could be a wide-sense stationary stochastic process with mean \( \bar{\nu} \), variance \( \sigma_v^2 \), and arbitrary autocorrelation function.) Let \( E(\cdot) \) denote the expectation operator taken over all possible realizations \( \{v_1, \ldots, v_T\} \), with the probability distribution of these determined by \( V \).

It follows that \( E(\cdot) = E[V(\cdot)] \). Therefore, the moments of \( y \) are found by computing the expectations of (2.1) and (2.3) with respect to \( V \),

\[
\begin{align*}
E(y_T^\gamma) & = E[V(y_T^\gamma)] = E(0) = 0 \\
E(y_T^{2\gamma}) & = E[V(y_T^{2\gamma})] = E(v_T) = \bar{\nu} \\
E(y_T^{4\gamma}) & = E[V(y_T^{4\gamma})] = \gamma E(v_T^{2\gamma}) = \gamma (\bar{\nu} + \sigma_v^2) \\
\end{align*}
\]

Similarly,

\[
E(\mu_1) = 0 \quad , \quad E(\mu_2) = \bar{\nu} \quad , \quad E(\mu_4) = \gamma (\bar{\nu} + \sigma_v^2) \quad (2.7)
\]

\[
\text{KURTOSIS}(y) = E[kurtosis}_V(y)] = \gamma (1 + c^2) \quad . \quad (2.8)
\]

The expected value of \( c^2 \) does not simplify because \( c^2 \) is the ratio of the first and second moments of the...
realization \( \{v_1, \ldots, v_T\} \). The distribution of this ratio depends in a complex fashion on the serial dependence in \( V \). However, the plim of the ratio will be the ratio of the corresponding population moments under appropriate assumptions concerning the moments of \( V \).

\[
\text{plim} \left( \text{KURTOSIS}(y) \right) = \gamma \left( 1 + \frac{\sigma_v^2}{\nu^2} \right)
\]  

(2.9)

If the autocorrelation of \( V \) is predominantly positive, the expected value of the coefficient of variation will be smaller than the population value for small sample sizes. (Heuristically, when \( v \) fluctuates slowly, the sample variance of a short series of successive values of \( v \) will be reduced, because a long sample is required to exhibit the full tendency of \( v \) to vary.) For this same reason, from (2.8) it follows that the population kurtosis of \( y \) will increase as the sample length grows. Finally, under assumptions on the existence of moments of \( F \) and \( V \), the plim of the sample kurtosis of \( y \) will again equal the population kurtosis.

\[
\text{plim} \left( \frac{\mu_4}{\mu_2^2} \right) = \gamma \left( 1 + \frac{\sigma_v^2}{\nu^2} \right)
\]  

(2.10)

2.3. It is also interesting to examine the behavior of moving sums of the \( y_t \) (which correspond, in the model of security prices, to the logarithms of price changes cumulated over several time intervals). Consider the sum of \( \ell \) successively observed \( y_t \), \( y_t^{\ell} = \sum_{s=t}^{t+\ell-1} y_s \). Let \( \nu_t^{\ell} = \sum_{s=t}^{t+\ell-1} v_s \). Assuming a fixed realization of the \( v_t \),
\[ E_y(\chi y_t) = 0 \]
\[ E_y(\chi y_t^2) = \chi v_t \]
\[ E_y(\chi y_t^4) = (\gamma - 3) \sum_{s=t}^{t+\lambda-1} v_s^2 + 3 \chi^2 v_t \]

\[ \text{KURTOSIS}_y(\chi y_t) = (\gamma - 3) \left( \frac{\sum_{s=t}^{t+\lambda-1} v_s^2}{\chi^2 v_t} \right) + 3. \]  

(2.11)

If successive values of \( v_s \) are similar, the first term in (2.12) will be similar to \( (\gamma - 3)/\chi \), and the kurtosis of the sum will converge toward 3 as \( \lambda \) increases.

The next inquiry is into the moments of the \( \chi y \) population, again assuming a fixed realization of the \( v_t \). Assume that the intervals of length \( \lambda \) over which the variables are summed are nonoverlapping, so that the population in question is the series of variables \( \chi y_t = \sum_{s=t}^{t+\lambda-1} y_s \), for \( t = 1, \lambda + 1, 2\lambda + 1, \ldots \), or, in more compact notation, the series of variables \( \chi y_{\lambda t + j}, \quad j = 0, \ldots, J - 1 \). For ease of comparison with the results for individual \( y_t \), assume that the length of the entire period is an integral multiple of \( \lambda \), so that \( T = J\lambda \). Define \( \chi y_i \equiv \sum_{j=0}^{J-1} \chi y_{\lambda j + 1} / J \) and \( \chi y_i \equiv \sum_{j=0}^{J-1} \chi y_{\lambda j + 1} / J \). Since the sums are taken over nonoverlapping intervals, the sets of \( \eta \) contributing to the different \( \chi y_t \) will be independent of one another. Therefore, 

---

---
\[ E_v(\xi_{11}) = 0 \quad , \quad E_v(\xi_{12}) = \sum_{j=0}^{J-1} \sum_{s=1}^{\ell} v_{j+s} \frac{1}{\ell} = \xi_{11} = \xi(\nu_1) \]

\[ E_v(\xi_{4}) = \frac{(\gamma - 3)}{J} \sum_{j=0}^{J-1} \sum_{s=1}^{\ell} v_{j+s}^2 \left( \frac{1}{\ell} \sum_{s=1}^{\ell} v_{j+s} \right)^2 = (\gamma - 3) \xi_{1} \nu_2 + 3 \xi \nu_2. \]

From these expressions, the population kurtosis can be derived.

\[ \text{KURTOSIS}_v(\xi y) = \frac{E_v(\xi_{4})}{(E_v(\xi_{12}))^2} = \frac{(\gamma - 3) \xi \nu_2 + 3 \xi \nu_2}{(\xi \nu_1)^2} \]

\[ = \frac{(\gamma - 3)}{\xi} \frac{\nu_2}{\nu_1^2} + 3 \left( \frac{\xi \nu_2}{\xi \nu_1} \right) = 3 + \left( \frac{\gamma - 3}{\xi} \right) (1 + \xi^2) + 3 \xi \xi^2 \]

(2.14)

where \( \xi \) is the coefficient of variation of the sample \( v_{j+l+1}, j=0, \ldots, J-1. \)

To facilitate comparison with the preceding case, (2.4) can be rewritten

\[ \text{KURTOSIS}_v(y) = 3 + (\gamma - 3)(1 + \xi^2) + 3 \xi^2. \]

(2.15)

Thus the deviation of the population kurtosis from normality is made up of two terms. The first, which reflects the nonnormality of \( \eta \), is multiplied by the factor \( (1 + \xi^2) \), but reduced by the factor \( 1/\xi \). The second term, the contribution of the variation of the \( \xi v \), occurs regardless of the kurtosis of the \( \eta \) distribution, and is proportional to \( \xi^2 \).

\[ \text{The inequality } \xi^2 \leq \xi^2 \text{ will always hold.} \]

\[ \rightarrow \text{At one extreme, if the } v_t \text{ are constant within each interval, so that all variation occurs between intervals, } \xi^2 = \xi^2. \text{ At another extreme, where successive values of } v_t \text{ within the same interval are unrelated (so that} \]

\[ \sum_{j=0}^{J-1} (v_{j+l+1} - \nu_1)(v_{j+l+2} - \nu_1) = 0 \text{ for all } n_1 < n_2 = l), \xi^2 = \xi^2/\xi. \text{ Where} \]

\[ J-1 \]

\[ \sum_{j=0}^{J-1} (v_{j+l+1} - \nu_1)(v_{j+l+2} - \nu_1) = 0 \text{ for all } n_1 < n_2 \leq l), \xi^2 = \xi^2/\xi. \text{ Where} \]

\[ j=0 \]
the $v_t$ follow a slowly changing stochastic process, with strong positive serial correlation over small intervals, the situation will lie somewhere between these two extremes and nearer the first. In this case, the contribution of the kurtosis of $\eta$ is reduced by the factor of $1/\lambda$, while the contribution from the variation of $v$ is reduced by the factor $c^2_\lambda / c^2 \sim 1$. On the other hand, if the sums of $\ell$ terms are chosen by random sampling, without replacement, from the series $\{y_1, y_T\}$, rather than as sequential sums, successive terms appearing in the same sum will be unrelated, so that the second of the two extremes will hold approximately, and

$$\text{KURTOSIS}_v(y) - 3 \sim \frac{\text{KURTOSIS}_v(y) - 3}{\lambda}.$$  \hspace{1cm} (2.16)

2.4. Finally, consider the sample moments of $\lambda v$ implied by the stochastic process $\nu$. As in Subsection 2.2., these moments are just the expected values, with respect to $\nu$, of the moments conditional on a fixed realization of $\nu$. Thus,

$$E(\lambda y_t) = 0$$

$$E(\lambda y_t^2) = \ell v$$

$$E(\lambda y_t^4) = \ell (v^2 + \sigma_v^2) + 3E[\lambda v_t^2]$$

$$\text{KURTOSIS}(\lambda y_t) = 3 + (\gamma - 3) E \left[ \left( \sum_{s=t}^{t+\ell-1} \frac{v_s^2}{\lambda v_t^2} \right) / \lambda v_t \right]$$  \hspace{1cm} (2.17)

$$E(\lambda u_1) = 0$$

$$E(\lambda u_2) = \lambda v$$

$$E(\lambda u_4) = \ell (v^2 + \sigma_v^2) + 3E[\lambda v_2]$$

$$\text{KURTOSIS}(\lambda y) = 3 + \frac{(\gamma - 3)}{\lambda} E(1+c^2) + 3E[\lambda c^2]$$
Where an expectation on the right-hand side has not been simplified, its value depends on the autocorrelation function of \( V \). Again under appropriate assumptions concerning the moments of \( F \) and of \( V \), the plims of these expectations will be the corresponding functions of the population moments of \( v \), and, moreover, the plims of the sample moments of \( \chi^2 \) will equal these population moments of \( \chi^2 \). Also, as before, the population kurtosis of \( \chi^2 \) will increase as sample size grows if \( v \) fluctuates slowly relative to the length \( \lambda \) of the differencing interval.

3. THE USE OF A REGRESSION MODEL TO ANALYZE FLUCTUATIONS IN VARIANCE

Consider briefly some of the factors which might influence the variance of price change in a security market. One factor is the elapsed time between successive price measurements. It is reasonable to assume that, other things being equal, the variance will be a linear function of both the number of hours of trading and the number of calendar days in the interval. However, the variance of price change will also depend on the variance of economic change per unit of time, and upon the amplitude of price response within each hour of trading. It is therefore helpful to divide other influences on price variance into two parts: the determinants of the variance of the economic, financial, social, and exogenous information reaching the market; and the determinants of the amplitude of the market's response to this information.

Within the first category are such factors as the variability of aggregate economic activity, the rate of technological innovation, the sensitivity of earnings to changes in cost or demand functions, the variance of exogenous changes in the asset positions of market participants, the
probability of significant changes in tax legislation, and the degree of sensitivity to international events. Within the second category are such factors as the trading mechanism within the market, the number and character of traders participating in the market, legal restrictions on traders' actions such as margin requirements and restrictions on short sales and options, the schedule of transactions costs, the vulnerability of the market to temporarily self-fulfilling but irrational expectations ("bullish" and "bearish" psychologies), and the traders' expectations as to the normal range of price fluctuations.

There will be a certain degree of feedback from the variance of market prices and from realized price changes to some of these factors, but all of them have an existence independent of contemporaneous events on the market floor. This is not true of the volume of transactions. It is reasonable that large price changes will tend to be accompanied by large volume, and vice versa, and significant relationships have been established between squared price changes and volume [7, Ch. 8]. However, the volume should not be treated as jointly determined with the variance, but rather as being jointly determined with the realized price change, for volume is generated in the trading which produces the realized price change. Factors that tend to increase the variance of price change will also tend to increase the expected volume, and vice versa. It is the relationship between these factors and the variance of price change that is of primary interest.

A relationship of this kind would afford the opportunity to forecast price variance and might also lead to constructive recommendations concerning market organization. Moreover, a description of historical fluctuations
in price variance would provide a basis for statistical analysis of the
historical stock price distribution, as well as a model of heteroscedasticity
in historical price changes that would be useful in a variety of other em-
pirical studies.

3.1. If the variance of market price change was an observable variable,
it would be possible to regress it directly upon explanatory variables repres-
senting the influences suggested above. In reality, the variance is not ob-
served, but the observed squared price change is a realization of the under-
lying distribution with expected value equal to the variance and, accord-
ingly, provides a means to carry out the regression. To clarify this, assume
that the true \( \sqrt{\text{determining the variances } v_t \text{ is of the form}} \)

\[
v_t = f(x_t, b, u_t) \quad , \quad t=1,\ldots,T \tag{3.1}
\]

where \( x_t \) is a \((k \times 1)\) vector of predetermined variables, \( b \) is a \((1 \times 1)\)
vector of parameters, and the disturbances \( u_t \) are serially independent,
identically distributed variables that are contemporaneously uncorrelated
with the predetermined variables. The variables \( \eta_t \), which were previously
assumed to be independent of the \( v_t \), are now assumed to be independent of
\( x_1,\ldots,x_K, b, \) and \( u_t \). Then the equation

\[
y_t^2 = f(x_t, b, u_t)\eta_t^2 \quad , \quad t=1,\ldots,T \tag{3.2}
\]

provides the basis for regression analysis.

For example, suppose that the variance is determined by the multipli-
cative model

\[
v_t = b_1 x_{1t} b_2 x_{2t} \cdots b_K x_{Kt} u_t \quad , \quad t=1,\ldots,T \tag{3.3}
\]
where the \( u_t \) have mean zero, variance \( \sigma^2 \), and are independent of \( x_1, \ldots, x_K \). Assume further that \( \eta_t \) is bounded away from zero, so that \( \log \eta_t^2 \) has a finite mean \( \alpha_1 \) and a finite variance \( \alpha_2 \). Then the logarithm of (3.2) is

\[
\log y_t^2 = \sum_{i=1}^{K} b_i \log x_{it} + u_t + \log \eta_t^2 \\
= \alpha_1 + \sum_{i=1}^{K} b_i \log x_{it} + \delta_t
\]

(3.4)

where \( \delta_t = u_t + \log \eta_t^2 - \alpha_1 \). This regression equation meets all of the conditions for the Gauss-Markov theorem, since the additive random variables \( \delta_t, \ t=1, \ldots, T \) are identically distributed with mean zero and variance \( \sigma^2 + \alpha_2 \) and are independent of \( x_1, \ldots, x_K \).

As another example, assume that the variance is determined by an additive model

\[
v_t = \sum_{i=1}^{K} x_{it} b_i + u_t = x_t' b + u_t, \quad t=1, \ldots, T
\]

(3.5)

where the disturbances \( u_t \) are again assumed to be serially independent with mean zero and variance \( \sigma^2 \). Then the regression equation may be written

\[
y_t^2 = v_t \eta_t^2 = v_t + v_t (\eta_t^2 - 1) \\
= x_t' b + u_t + (x_t' b + u_t)(\eta_t^2 - 1) \\
= x_t' b + \varepsilon_t, \text{ where } \varepsilon_t = u_t + (x_t' b + u_t)(\eta_t^2 - 1).
\]

(3.6)

The stochastic properties of the series \( \varepsilon_t, \ t=1, \ldots, T \), are not ideal, but the least squares estimator based on (3.6) does provide a consistent estimator
of \( b \). To demonstrate this, the relevant moments of the \( \varepsilon_t \), taken with respect to the specification of the stochastic process \( V \), must be evaluated. In consistency with the notation of the previous section, let 
\[ E(\cdot) = E(\cdot | x_1, \ldots, x_k, b, \sigma^2). \]
Then
\[ E(\varepsilon_t) = E(u_t) + E(x_t'b + u_t)E(\eta_t^2 - 1) = 0 \]  
(3.7)
since the components of \( v_t \) are independent of \( \eta_t \), and since \( E(u_t) = 0 \) and \( E(\eta_t^2) = \text{VAR}(\eta_t) = 1 \). Similarly,
\[ \text{VAR}(\varepsilon_t) = E(\varepsilon_t^2) = E[u_t^2] + 2E(u_t (x_t'b + u_t))E(\eta_t^2 - 1) + E[(x_t'b + u_t)^2]E[(\eta_t^2 - 1)^2] \]
\[ = \sigma^2 + (x_t'b)^2 + \sigma^2(\gamma - 1). \]  
(3.8)
Also, \( E(x_t \varepsilon_t) = 0 \), because of the zero correlation between \( u_t \) and \( x_t \), and the independence between \( \eta_t \) and \( x_t \). Therefore, with the mild assumption that \( \text{plim} \left( \frac{\sum_{t=1}^{T} x_t x_t' / T}{T} \right) \) exists and is nonsingular, the probability limit of the least squares estimator \( \hat{b} = \left( \sum_{t=1}^{T} x_t x_t' \right)^{-1} \sum_{t=1}^{T} x_t y_t \) is
\[ \text{plim} \hat{b} = \text{plim} \left( \frac{\sum_{t=1}^{T} x_t x_t'}{T} \right)^{-1} \left( \sum_{t=1}^{T} x_t(x_t'b + u_t) + (x_t'b + u_t)(\eta_t^2 - 1) \right) \]
\[ = b + \text{plim} \left( \frac{\sum_{t=1}^{T} x_t x_t}{T} \right)^{-1} \left( \frac{\sum_{t=1}^{T} x_t u_t}{T} + \frac{\sum_{t=1}^{T} x_t(x_t'b + u_t)(\eta_t^2 - 1)}{T} \right) = b. \]  
(3.9)
Thus, the random terms \( \varepsilon_t \) are heteroscedastic and dependent upon the explanatory variables and regression parameters, but the dependence is limited
to such a degree that least square regression provides consistent estimators of the regression parameters.

3.2. The results of Section 2 concerning the kurtosis of \( y \) can be easily applied to a stochastic process \( V \) of form (3.1). The details for the linear regression model (3.5) will be worked out here. Let \( \overline{z} = \frac{1}{T} \sum_{t=1}^{T} x_t / T \) and let \( H = \sum_{t=1}^{T} (x_t - \overline{z})(x_t - \overline{z})' / T \). Then the moments of the realization \( \{v_1, \ldots, v_T\} \) are

\[
\overline{v} = E[v_1] = \overline{x}'b
\]

(3.10)

\[
\overline{v}^2 + \sigma_v^2 = E[v_2] = (\overline{x}'b)^2 + b'HB + \sigma^2
\]

(3.11)

When these expressions are substituted into (2.9), the limiting value for the kurtosis of \( y \) is found to be

\[
\gamma_y \equiv \text{plim}[\text{KURTOSIS}(y)] = \gamma \left( 1 + \frac{b'HB + \sigma^2}{(\overline{x}'b)^2} \right)
\]

(3.12)

The moments of the moving sums of the \( y_t \) can be found similarly by substituting (3.10) and (3.11) into (2.17). The terms in (2.17) that depend on the serial dependence of the \( v_t \) are now determined by the behavior over time of the deterministic component \( \overline{x}'b \).

Equation (3.12) provides consistent means of estimating \( \gamma \). With appropriate assumptions on the existence of higher moments of \( u \) and \( \eta \), the sample kurtosis of \( y \) provides a consistent estimator for \( \gamma_y \), the mean of \( y_t^2, t=1, T \), provides a consistent estimator for \( \overline{v} \), and the estimator \( \hat{b} \) provides a consistent estimator for \( \hat{b}'H \), namely \( \hat{b}'H \).
Unfortunately, estimation of $\sigma^2$ is somewhat problematic. The difficulty is to distinguish the contribution of $\sigma^2$ to the residual variance in the regression (3.6) from the contribution of $\gamma$. Examining (3.8), it is apparent that these two contributions can, in principle, be discriminated, for the error variance contributed by $\sigma^2$ is identical for all observations, whereas the variance contributed by $\gamma$ is a linear function of $x_t'b$. Thus, using a consistent estimator of $b$, the linear regression

$$\left(\gamma - x_t'b\right)^2 = \gamma \sigma^2 + \left(x_t'b\right)^2(\gamma - 1) + \zeta_t, \quad t=1,\ldots,T, \tag{3.13}$$

where the left-hand variable is the squared residual in the fitted regression (3.6), will yield consistent estimators of $\gamma$ and $\sigma^2$, if the appropriate higher moments of $u$ and $\eta$ exist. However, this approach is not robust against minor misspecifications in the model. Also, since the dependent variable involves the fourth power of the price change, the outcome will be extremely sensitive to the values of a few realized price changes. Rather than relying upon this approach or upon an analogous but more sophisticated device, it may be preferable to take the conservative stance that $\sigma^2$ is zero, which is equivalent to assuming that $\gamma$ is maximal. This approach yields a consistent upper bound for $\gamma$,

$$\gamma = \gamma \left(\frac{-\hat{\sigma}^2}{\hat{\sigma}^2 + b'\hat{H}\bar{b} + \hat{\sigma}^2} \right) \leq \gamma \left(\frac{-\hat{\sigma}^2}{\hat{\sigma}^2 + b'\hat{H}\bar{b}} \right) \leq \gamma \left(\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + b'\hat{H}\bar{b}} \right). \tag{3.14}$$

To conclude this subsection, note that the consistent estimators for $\gamma$ and $\sigma^2$ provide an estimate of the heteroscedasticity in the residuals in the basic regression (3.6). Aitken's Generalized Least Squares can
therefore be applied to (3.6), using the estimated diagonal variance-covariance matrix (3.8), to yield asymptotically more efficient estimators of $\beta$.

3.3. The remaining task of this section is to examine an interim method of analyzing the variances of security price changes that can be used in the absence of operational measurements of the exogenous factors influencing this variance. The suggested approach is to employ a moving average of realized squared price changes as an estimate of the prevailing variance. The two moving-average regression models which will be used in the experiment, each of which will be estimated separately for values of $\lambda$ from 1 through 6, are:

$$\lambda y_t^2 = \alpha + \beta A_t$$

(3.15)

where

$$A_t = \frac{\sum_{s=2}^{11} y_{t-s}^2}{10}$$

and

$$\lambda y_t^2 = \alpha + \beta B_t$$

(3.16)

where

$$B_t = \frac{\sum_{s=2}^{11} (y_{t-s}^2 + y_{t+(\lambda-1)+s}^2)/20.}$$

The moving-average approach is a natural one that could stand on its own without any reference to the more general regression model proposed previously. However, the broader perspective in which the variances are viewed as the consequence of identifiable underlying factors allows a much richer interpretation of the results of the moving average model. Suppose that the true model for $v$ is of the form

$$v_t = x_t' b + \sum_{s=1}^{S} w_s v_{t-s} + \sum_{r=1}^{R} \omega_r y_{t-r}^2 + u_t$$

(3.17)

$$t=1, \ldots, T.$$
As before, the \( x \) variables represent exogenous factors. In addition, an autoregression on lagged values of the variance is included, since the \( \mathcal{Y} \) process may exhibit inertia or some other internal dynamics. Moreover, a distributed lag on the realized values of the squared price changes is included, since the dispersion of the price changes actually observed in preceding periods may influence traders' expectations as to the magnitude of price change in the present period. Equation (3.17) yields the regression relation

\[
y_t^2 = x_t' b + \sum_{s=1}^{S} \sigma_{s} v_{t-s} + \sum_{r=1}^{R} \omega_{r} y_{t-r}^{2} + \epsilon_t, \quad t-1, \ldots, T. \tag{3.18}
\]

If this is the true regression model, the moving-average regressions are misspecified, and the results must be interpreted in this light. The contributions of each of the three kinds of predetermined variables in (3.17)—the exogenous factors, the lagged variances, and the lagged squared price changes—will be captured in the regression coefficients for the moving average only to the degree that the moving average is a surrogate for (is correlated with) these variables. In discussing the implications of the misspecification, it is sufficient to touch on three important points that are crucial to the interpretation of the results.

First, the moving averages can capture the effects of the exogenous factors only to the degree that these factors are slowly changing over time. Many of the important factors cited earlier in this section do change slowly relative to the one-month recording intervals in the data. However, some factors change too quickly to be captured. In particular, the duration of the month, both in terms of calendar days and in terms of hours of trading, will fluctuate from month to month in response to the calendar and cannot be
captured by the moving average. Fortunately, this factor is of minor importance in studying monthly data; the squared coefficient of variation is less than .01, so that from (2.8) the contribution to the kurtosis of $y$ is, at most, 1 percent.

Second, there is a problem of simultaneity in using the two-sided moving average as an explanatory variable. The average includes future observations, $y_{t+s}^2$, which depend on future values, $v_{t+s}$. But according to (3.17), $v_{t+s}$ depends on the current value of $y_t^2$, both directly through the distributed lag and indirectly through the autoregression on $v_t$. Hence, the regression coefficient for the two-sided moving average will capture not only some of the determinants of $v_t$, but also part of the reverse effect of $v_t$ and $y_t^2$ on future values of $v$. This problem does not arise with the lagged moving average.

Third, in using the lagged moving average of $y^2$ as a surrogate for the exogenous factors and for the autoregression on $v$, an interesting "errors in variables" problem arises. To clarify this issue, assume that the exogenous factors change so slowly that $x_t'b$ is highly correlated with $v_t'$, thus permitting these two terms to be merged. Assume further that the weights in the distributed lags in the true model correspond to the lagged moving average so that the true model becomes

$$y_t^2 = \alpha + \phi v_t + \omega A_t + \epsilon_t \quad , \quad t=1, \ldots , T, \quad (3.21)$$

where $V_t = \frac{11}{\Sigma v_{t-s}/10}$ and $A_t$ is defined as before. Hence, in the regression model (3.15), $A_t$ serves, in part, to capture its own effect and, in part, as a surrogate for $V_t$. But $A_t$ is related to $V_t$, for
\[ A_t = \frac{11}{10} \sum_{s=2}^{11} y_{t-s}^2 / 10 = \frac{11}{10} \sum_{s=2}^{11} (v_{t-s} + \varepsilon_{t-s}) / 10 = v_t + \theta_t, \] (3.22)

where \( \theta_t = \frac{1}{10} \sum_{s=2}^{11} \varepsilon_{t-s} / 10. \)

It is useful to distinguish two extreme cases. At one extreme, where \( \phi = 0, \) (3.21) coincides with (3.15), the regression model is correctly specified, and the results can accordingly be accepted at face value. At the other extreme, where \( \omega = 0, \) the true model is

\[ y_t^2 = \alpha + \phi v_t + \varepsilon_t, \] (3.23)

and (3.15), (3.22), and (3.23) constitute a pure case of the errors-in-variables problem. This is an unusually interesting case, because the appropriate corrections to achieve consistent estimators can be derived. From the specification of the model,

\[ E[\varepsilon_t] = E[\varepsilon_t] = E[\theta_t \varepsilon_t] = E[\theta_t v_t] = E[\varepsilon_t v_t] = 0 \] (3.24)

\[ \text{VAR}[\varepsilon_t] = (\gamma - 1)v_t + \sigma^2, \quad \text{VAR}[\theta_t] = \sum_{s=2}^{11} \text{VAR}[\varepsilon_{t-s}]/100. \]

On the assumption that the \( v_t \) change slowly over time, \( \text{VAR}(\theta_t) = \text{VAR}(\varepsilon_t)/10, \) and the results will be little effected if it is assumed that for all \( t \) the individual variances are equal to their average values, \( \text{VAR}[\varepsilon_t] = \text{VAR}(\varepsilon), \)

\[ \text{VAR}(\phi_t) = \text{VAR}(\varepsilon)/10. \] Then applying the familiar approach to the errors in variables problem (e.g., [9, pp. 281-291]), it follows that

\[ \text{plim}(\hat{\beta}) = \phi \frac{\text{VAR}(V)}{\text{VAR}(V) + \text{VAR}(\varepsilon)/10} \] (3.25)

Also, from (3.23),
\[ \text{VAR}(y^2) = \phi^2 \text{VAR}(V) + \text{VAR}(\epsilon). \]  \hfill (3.26)

Finally,
\[ \text{VAR}(A) = \text{VAR}(V) + \text{VAR}(\epsilon)/10. \]  \hfill (3.27)

Since \( \hat{\phi}, \text{VAR}(y^2), \) and \( \text{VAR}(A) \) are computed statistics, these three equations in the three unknowns \( \phi, \text{VAR}(V), \) and \( \text{VAR}(\epsilon) \) may be solved to obtain estimators of \( \hat{\phi} \) and \( \text{VAR}(V). \) More generally, if the dependent variable in the regression is the square of the price change cumulated over \( k \) periods, so that the assumed true model is
\[ y_t^2 = \alpha + \phi V_t + \sum_{s=0}^{k-1} \epsilon_{t+s}, \]  \hfill (3.28)

(3.25) and (3.27) hold as before and (3.26) is modified to
\[ \text{VAR}(y^2) = \phi^2 \text{VAR}(V) + k \text{VAR}(\epsilon). \]  \hfill (3.26*)

The solutions of (3.25), (3.26*), and (3.27) yield an upper bound for \( \gamma, \) which has been corrected for the errors-in-variables effect,
\[ \gamma \leq \gamma_y \left( \frac{\text{VAR}(V)}{\text{VAR}(y^2) + \sigma^2 \text{VAR}(V)} \right) = \gamma_y \left( \frac{\gamma_y^2}{\gamma_y^2 + \phi^2 \text{VAR}(V)} \right). \]  \hfill (3.29)

In summary, if the variance of price change is determined by a distributed lag on realized squared price changes, then the approximate upper bound is (3.14), which simplifies in the special case (3.15) to
\[ \gamma \leq \gamma_y \left( \frac{\text{VAR}(V)}{\text{VAR}(y^2) + \sigma^2 \text{VAR}(V)} \right) = \gamma_y \left( \frac{\gamma_y^2}{\gamma_y^2 + \phi^2 \text{VAR}(V)} \right). \]  \hfill (3.30)

On the other hand, if the variance of price change is determined by an autoregression perturbed by exogenous factors, the upper bound (3.29) is appropriate.
If the truth lies somewhere in between, then the correct upper bound will lie between these extremes.

4. THE DATA

The price series used in the experiment is the Standard & Poor's Monthly Composite Stock Price Index (the S&P 500). The values of this index and of S&P indexes for many industrial subgroups are readily available for the period from 1871 to the present [18]. The Composite Index is a natural choice for a basic experiment, not only because of its extraordinarily long history, but also because it closely parallels the total value of common stock equity in the United States.

4.1. Under the null hypothesis stated in Section 1, the variance of the random increment in the logarithm of price is proportional to the elapsed time over which the increment occurs. Therefore, since the average month contains 4-1/3 weeks, the variance of $\ell Z^w_{\ell}$ is approximately $4.333 \ell \nu_w$, where $\nu_w$ is the variance of the increment between successive Wednesdays. Also, the series of differences of logarithms taken over successive nonoverlapping intervals are serially independent with variances that will be equal except for the minor variations due to the varying number of days in a month.

There are more than 100 years, or 1200 months, of data. Thus, there are 1200+ observations for the case $\ell = 1$, 600+ for $\ell = 2$, 400+ for $\ell = 3$, 300+ for $\ell = 4$, 240+ for $\ell = 5$, and 200+ for $\ell = 6$. To insure that the sample size remains large, the maximum differencing interval used is $\ell = 6$. 
(The apparent sample size could be increased by including all 1200 possible differencing intervals of length \( l \) in the same sample, or by studying \( l \) distinct but overlapping samples, \( i=1, \ldots, l \), each made up of the series of nonoverlapping intervals starting with the \( i \)th month in the sample. However, the sampling properties of the results are clearer when nonoverlapping intervals are used, and there are sufficient data in the sample to prove the point of the article in this way.)

Let \( C_t \) denote the value of the composite index in month \( t \), and let \( \xi_t = \log C_{t+\ell} - \log C_t \). It would appear to be entirely natural to use each variable \( \xi_t \) to operationalize the corresponding variable \( \xi^2_t \), since this amounts to using the value of the composite price index as a measure of the price in that month. If the composite index were an instantaneous measure of the level of market prices at some fixed time in the month, \( \xi_t \) would indeed be a measure of the movement in the logarithm of "market price" from one instantaneous recording moment to another \( \ell \) months later, and it would be entirely appropriate to use it as the variable \( \xi^2_t \).

Unfortunately, the Standard & Poor's indexes are not constructed so as to be instantaneous measures of the market price level. The Index is actually a splice of two differently constructed indexes: until December 1917, the monthly price index was a weighted average of individual security "prices," with each price, because of data limitations, defined as the monthly mid-range (the average of the high and low in the month); since January 1918, the monthly price index has been defined as the average, taken over all Wednesdays in the month, of the weighted average of the individual securities' Wednesday closing prices. The reasons for this change in definition and the exact nature of the indexes are explained in detail in [3]. Each definition
yields a "monthly price" which is actually the result of applying some averaging measure to the entire monthly price history. The averaging measures alter the variances of the price changes and destroy the property of serial independence as well. This point is demonstrated elegantly in [4] and further discussed in [17]. Following the approaches taken in [4], [16], and [17], the extent of the reduction in variance and the magnitude of the resultant serial correlation can be deduced for each of the averaging measures.

For the mid-range, with the assumption that the null hypothesis holds and, in addition, the assumption that \( z_t \) is normally distributed and some simplifying assumptions about the procedure for constructing the mid-range index,

\[
\text{VAR}(\chi c_t) = ((\ell-1)4.333+2.659)\nu_w = \frac{(\ell-1)+.6137}{\ell} \text{VAR}(\chi z_t).
\]

(4.1)

For the post 1918 period, following the approach sketched in [17] and with attention to the varying numbers of Wednesdays in successive months,

\[
\text{VAR}(\chi c_t) = ((\ell-1)4.333+2.967)\nu_w = \frac{(\ell-1)+.6846}{\ell} \text{VAR}(\chi z_t).
\]

(4.2)

The spurious shift in the variance of \( c_t \) in January 1918 would be taken as a minor validation of the alternative hypothesis if it were not corrected for. Accordingly, the series of log differences prior to 1917 must be corrected by a multiplicative factor so that, under the null hypothesis, the corrected pre-1917 and post-1918 variances are identical. If the effects of the two averaging measures are exactly as asserted in (4.1) and (4.2), the factor should be

\[
f_\ell = \sqrt{((\ell-1) + .6846)/((\ell-1) + .6137)}
\]

for differencing intervals of length \( \ell \).
4.2. Next consider the possibility of dependence between two price changes, \( \xi^c_s \) and \( m^c_t \), for \( s < t \). As long as the first interval ends in a month prior to the month in which the second interval begins, the serial independence of the increments in the underlying price process assures that the price changes will be independent of one another. However, if the first interval ends in the same month as the second begins, the two measured changes each involve the same averaging measure \( C_t \) taken over the entire month \( t \), and, accordingly, serial dependence enters in. For the case of successive differencing intervals of length \( \lambda \), the spurious serial correlation can be shown to be equal, on the average, to

\[
\text{CORR}(\xi^c_{t-\lambda}, \xi^c_t) = (1-a)/2((\lambda-1) + a),
\]

(4.3)

where \( a = \text{VAR}(1_c)/\text{VAR}(1_z) \).

In particular, the spurious correlation of successive first differences equals .315 for the pre-1917 period, and .230 for the post-1918 period, if the variance reductions are as given in (4.1) and (4.2).

Relation (4.3) provides a check of the validity of (4.1) and (4.2), for the spurious reduction in variance is uniquely related to the spurious serial correlation introduced into the series. For the pre-1917 period (558 observations), the actual serial correlation of the series \( 1^c_t \) is .315, exactly equal to the predicted spurious serial correlation. For the post-1918 period (635 observations), the actual serial correlation is .291, in contrast to the predicted spurious serial correlation of .230. This difference of .061 is marginally significant, since the asymptotic approximation to the standard error of the estimated correlation coefficient is .036 under the null hypothesis [10, III, p. 432]. It may reflect a somewhat larger variance reduction than predicted in (4.2), or positive serial correlation in the increments to
the underlying price process, or any combination of these two causes. In fact, the cause may be serial correlation in the underlying price changes, for .061 is in the same range as serial correlations observed for the first differences of various price series over short differencing intervals in the period 1920-1970. In sum, the observed serial correlation of the composite index appears to be largely consistent with the predictions of (4.3), and therefore appears to confirm (4.1) and (4.2).

However, to assure that misspecification of the correction factor did not affect the results in any way, the analyses were run for three cases: no correction factor; the correction factor \( f_\lambda \); and the correction factor lying between these cases, which would be appropriate if all of the serial correlation in the post-1918 period were spurious. The results never differed importantly from one case to the next, so only those results with the correction factor \( f_\lambda \), which are the least favorable to the alternative hypothesis, will be reported.

To facilitate the interpretation of the results, each corrected series \( \lambda^c_t \) was also multiplied by a constant multiplicative factor \( h_\lambda = \sqrt{\lambda/((\lambda-1) + .6846)} \) over the whole history. This correction offsets the variance reductions in (4.2), so that under the null hypothesis, the transformed variance is equal to the variance of the underlying \( \lambda^z_t \). The transformation of scale in no way affects the explanatory power of the regression.

Some additional minor adjustments to the data were needed to deal with the closure of the New York Stock Exchange for four months in 1914—differencing intervals overlapping this period were deleted—and to deal with the discontinuous change in the variance reduction in January 1918. The series \( \lambda^c \) as corrected were then used in the regressions.
4.3. Two problems arise from the spurious serial dependence in the is $c_t$ (which / not removed by the multiplicative corrections). First, it is desirable that each dependent variable $\lambda_t$ be independent, under the null hypothesis, of the moving average used to predict its variance. The averaging measures render $\lambda_t$ dependent on any price changes which involve prices in month $t$ through $t + \ell$. However, $\lambda_t$ remains independent of any of the $1^c_s$ taken over differencing intervals that end prior to $t$ or begin subsequent to $t + \ell$. Accordingly, the moving averages are defined so as to omit the variables $1^c_{t-1}$ through $1^c_{t+\ell}$ inclusive, thereby insuring the desired independence under the null hypothesis. Second, the spurious first-order serial correlation of the $\lambda_t$ induces a spurious serial correlation in successively observed $\lambda_t^2$ and, hence, in the residuals in the regressions (3.15) and (3.16). However, it can be shown that the spurious serial correlation in the regression residuals will be less than $.1/\ell$, so that there will be a negligible reduction in the efficiency of the least squares estimators.

One last problem is the effect of the averaging measures upon the kurtosis of the price changes. For the pre-1917 period, where the mid-range was used, if $1z_t$ is normally distributed with kurtosis of 3, the kurtosis of $1^c_t$ is reduced to 2.798 under the null hypothesis [16]. Anticipating the results of the experiment, which imply that the distribution of $z_t$ is similar to the normal, it may be hoped that the reduction in kurtosis was of this magnitude. The effect on the kurtosis of the post-1918 averaging measure can be computed directly under the null hypothesis. The effect turns out to be small: $\text{KURTOSIS}(1^c_t) = .95(\text{KURTOSIS}(1z_t) - 3) + 3$. Overall, the kurtosis of $1^c_t$ can be expected to be similar to that of $1z_t$, but slightly lower.
Whatever difference there is between the two kurtoses decreases sharply as \( l \) increases. In summary, the success of the experiment in explaining the kurtosis of the corrected changes in the S&P Composite Index will be strongly indicative of the success that could be achieved in explaining the kurtosis of the underlying price changes, but the two results would not be identical.

5. THE EXPERIMENT

The first step in studying the data is to analyze the mean values of the price changes. The mean value constitutes one component of expected return, so it should be relatively small in magnitude and should change slowly over time, but there is no reason to expect it to be zero. However, polynomial distributed lag regressions applied to the series \( l_{ct} \) (not yet corrected as explained in Section 4, since the multiplicative correction factor perturbs the mean if this is nonzero) did not uncover any systematic patterns. Even the mean value of the series, computed separately over the periods 1871-1917 and 1918-1971, is not significantly different from zero at the 99 percent level of confidence in either period. The explanation is not that the true mean is zero, but rather that over differencing intervals of a few months the variance of stock price changes entirely swamps the contribution of the mean. This finding is consistent with other experience [5, p. 400], all of which seems to show that adjustments for the means of stock price changes have a negligible effect in studies of their probability distributions. Accordingly, rather than make a problematic assumption about the pattern of the mean values over time, the means of the \( z_t \) were assumed to be identically zero for all \( t \). Under this assumption, the corrected series \( l_{ct}^c, l = 1, 6 \) will take the role of the series \( l_{yt}^v \) as defined in Sections 2 and 3.
It was decided to divide the sample into two parts: one, covering the years 1871-1950, to be used in the regressions; and the other, covering the years 1951-1971, to be reserved to test the predictive ability of the models. When the experiment was begun, the length of the moving average to be used in the regressions had not yet been determined. Lagged moving averages of lengths 10 and 30 were tried, so that the first 32 months of the sample were sacrificed to initialize the explanatory variables. The moving average of length 10 produced a better fit and was therefore selected for the experiment. (No further effort was made to select the length of the moving average so as to maximize goodness of fit, since a search of this kind might cast some doubt on the eventual results.)

Then regressions of the form (3.15) and (3.16) were run. The results are reported in Tables 1 and 2. Since the series of dependent variables for each \( \ell \) is based on the same series of underlying price changes, the regression results are jointly dependent. Accordingly, in comparing the null hypothesis with the alternative hypothesis, it is appropriate to rely only upon one of the regressions—specifically, the regression for \( \ell = 1 \), with the largest number of observations and, hence, the largest amount of information.

\[ \beta \]

Under the null hypothesis, the regression coefficient \( \beta \) is equal to zero in all regressions. However, for \( \ell = 1 \), the t-statistic for \( \beta \) in the regression on the lagged moving average is 10.06. The 99.99 percent confidence point for the t-distribution is 4.1, so that even allowing for inflation in the tails of the distribution of the t-statistic as a result of heteroscedasticity and high kurtosis in the regression disturbances, the experiment appears to provide conclusive evidence for the existence of forecastable fluctuations in variance. Similarly, the t-statistic for \( \beta \) in the
TABLE 1

REGRESSIONS OF THE FORM \[ \sum_{t-s}^{t} \gamma^2 = \alpha + \beta \Delta_e, \] WHERE \[ A_t = \sum_{s=2}^{11} \gamma^2 t-s / 10 \]

FOR NONOVERLAPPING INTERVALS OF \( \ell \) MONTHS
FROM THE SAMPLE PERIOD AUGUST 1873–DECEMBER 1950

<table>
<thead>
<tr>
<th>( \ell )</th>
<th>Sample Size</th>
<th>Moments of ( \gamma^2 )</th>
<th>( R^2 )</th>
<th>( \hat{\alpha} ) ( \hat{\beta} ) (t-Statistics)</th>
<th>( \hat{\phi} )</th>
<th>Upper Bounds on ( \gamma ) From Eq. (3.30) From Eq. (3.29)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>923</td>
<td>.31</td>
<td>1.31</td>
<td>.099</td>
<td>.00106 (.257) .666 (10.06)</td>
<td>1.067</td>
</tr>
<tr>
<td>2</td>
<td>461</td>
<td>.66</td>
<td>4.69</td>
<td>.154</td>
<td>.00167 (1.55) 1.578 (9.14)</td>
<td>3.421</td>
</tr>
<tr>
<td>3</td>
<td>307</td>
<td>.92</td>
<td>4.80</td>
<td>.181</td>
<td>.00406 (3.12) 1.637 (8.21)</td>
<td>2.540</td>
</tr>
<tr>
<td>4</td>
<td>230</td>
<td>1.15</td>
<td>5.58</td>
<td>.103</td>
<td>.00693 (4.01) 1.477 (5.13)</td>
<td>2.596</td>
</tr>
<tr>
<td>5</td>
<td>184</td>
<td>1.60</td>
<td>9.79</td>
<td>.119</td>
<td>.00961 (3.80) 2.038 (4.85)</td>
<td>4.291</td>
</tr>
<tr>
<td>6</td>
<td>153</td>
<td>1.96</td>
<td>18.10</td>
<td>.134</td>
<td>.01078 (2.91) 2.822 (4.84)</td>
<td>7.615</td>
</tr>
</tbody>
</table>
TABLE 2

REGRESSIONS OF THE FORM $\ell y_t^2 = \alpha + \beta \ell^B_t$, WHERE

$$\ell^B_t = \frac{11}{20} \sum_{s=2}^{11} \left( y_{t-s}^2 + y_{t+(\ell-1)+s}^2 \right)$$

FOR NONOVERLAPPING INTERVALS OF LENGTH $\ell$ FROM THE SAMPLE PERIOD AUGUST 1873–DECEMBER 1950

<table>
<thead>
<tr>
<th>$\ell$</th>
<th>Sample Size</th>
<th>$R^2$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(t-Statistics)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>923</td>
<td>.1360</td>
<td>.00030 (0.70)</td>
<td>.913 (12.04)</td>
</tr>
<tr>
<td>2</td>
<td>461</td>
<td>.1955</td>
<td>.00006 (0.06)</td>
<td>2.098 (10.56)</td>
</tr>
<tr>
<td>3</td>
<td>307</td>
<td>.2553</td>
<td>.00185 (1.42)</td>
<td>2.339 (10.22)</td>
</tr>
<tr>
<td>4</td>
<td>230</td>
<td>.3590</td>
<td>.00144 (0.94)</td>
<td>3.260 (11.30)</td>
</tr>
<tr>
<td>5</td>
<td>184</td>
<td>.3895</td>
<td>.00230 (1.04)</td>
<td>4.392 (10.77)</td>
</tr>
<tr>
<td>6</td>
<td>153</td>
<td>.5204</td>
<td>-.00168 (-0.57)</td>
<td>6.813 (12.79)</td>
</tr>
</tbody>
</table>
regression on the two-sided moving average is 12.04, providing yet more conclusive evidence for the existence of systematic fluctuations in variance.

The estimated \( \hat{\beta} \) are also highly significant for \( \lambda > 1 \). The value of \( \hat{\beta} \) should increase in direct proportion to \( \lambda \), since the variance of price change should be the sum of its components. However, in regressions on the lagged moving average, the increase is less than in proportion to \( \lambda \). One might first suspect the corrections designed to introduce comparability across different \( \lambda \) (explained at the end of Section 4.2), but the means of the squared price changes given in the third column of Table 1, which are unbiased estimates of the average variances, do increase in proportion to \( \lambda \), so that these corrections are seen to have been successful. Since the adjusted regression coefficients \( \hat{\phi} \) are more closely proportional to \( \lambda \), it appears that a part of the difficulty is removed by the errors-in-variables approach.

The remaining discrepancies, which show up in the \( \hat{\phi} \) and also in the \( \hat{\beta} \) estimated for the two-sided moving average, are probably due to the variations in the sample variances of the \( x^2 \) as \( \lambda \) increases, seen in column 4 of Table 1, which are accidental consequences of the timing of the differencing intervals for each \( \lambda \).

At the right-hand side of Table 1, consistent upper bounds on \( \gamma \) (the kurtosis of the individual price changes), are computed for the two cases described in Section 3.3. Under the assumption that variance is determined by a distributed lag on realized squared price changes (3.30), the upper bounds lie somewhat above the kurtosis of the normal distribution. Under the assumption that the variance is determined by an autoregression perturbed by exogenous factors (3.29), the upper bounds fluctuate below the kurtosis of the
normal distribution. The arguments of Section 3 suggest that the true upper bound should lie between these two extremes.

The estimated regressions provide rules for predicting the variances of the historically observed price changes, respectively, \( \hat{v}_t = \alpha + \beta A_t \) and \( \hat{v}_t = \alpha + \beta B_t \). Each series \( \hat{v}_1, \ldots, \hat{v}_N \) can be standardized so as to have geometric average equal to unity. Let \( \hat{s}_t^2 \) denote such a standardized variance prediction. Consider the contrast between the series of variables \( \hat{v}_1, \ldots, \hat{v}_N \) and the adjusted series \( \hat{v}_1 / \hat{s}_1, \ldots, \hat{v}_N / \hat{s}_N \). Under the null hypothesis, the former are identically distributed, while the latter are heteroscedastic as a result of the erroneous division by \( \hat{s}_t^2 \). In contrast, under the alternative hypothesis, the former are heteroscedastic, while the latter should be less heteroscedastic as a result of division by the predicted standard deviations. Thus, by comparing the properties of the two series, additional insight into the relative validity of the two hypotheses is obtained.

For one thing, since the predicted variances are standardized to have geometric average equal to one, any reduction in the dispersion of \( \hat{v} \) achieved by the adjustment should reflect a true explanation of heteroscedasticity. A substantial reduction is achieved: for regression on the lagged moving average, \( \Sigma (v_t / \hat{s}_t)^2 / \Sigma v_t^2 = .802 \); for regression on the two-sided moving average, \( \Sigma (v_t / \hat{s}_t)^2 / \Sigma v_t^2 = .757 \).

Moreover, if the alternative hypothesis holds, division by the predicted standard deviations should remove fluctuations in variance and thereby reduce the kurtosis of the population. Conversely, if the null hypothesis holds, \( 1 / \hat{s}_t^2 \) and \( v_t \) are independent except for the
tenuous link provided by the two estimated regression coefficients. Therefore, according to the results of Section 2.1, division by the predicted standard deviations should increase the kurtosis. In Figure B, the histogram of the adjusted series \( \frac{Y_t}{\hat{s}_t} \), where \( \hat{s}_t \) is predicted on the basis of the lagged moving average, is superimposed on the same normal distribution as in Figure A. Comparing the two figures, the reduction in both scale and kurtosis achieved by the adjustment is apparent. A quantitative indication of the reduction in kurtosis is given in Tables 3 and 4. The reduction in kurtosis over the entire sample history, using the regression coefficients estimated from the regression subsample only, is very great.

However, the change in kurtosis taken over the subsample reserved for predictive testing is disappointing. The prediction based on the lagged moving average reduces the kurtosis in only one case out of six, the prediction based on the two-sided moving average reduces it in four cases out of six. The explanation for this mediocre performance is easily seen in Figure C, where the standardized series \( Y_t \) is plotted against a standard deviation band predicted by the lagged moving average. It is clear that the variance fluctuated greatly in the period 1920-1950, but very little in the period 1951-1971. Thus, when the lagged moving average is applied over the entire sample, there are major fluctuations in variance to explain; the sample kurtosis, which is originally very high as a result of these fluctuations, is reduced substantially when they are successfully predicted by the regression. In the reserved subsample, however, there is little fluctuation in variance; as a result, the sample kurtosis is not very high originally, and the corrections implied by the regression model introduce enough noise to offset largely the
C. STANDARDIZED MONTHLY PRICE CHANGES, 1872-1971, SUPERIMPOSED UPON A BAND OF WIDTH EQUAL TO TWO FORECAST STANDARD DEVIATIONS
TABLE 3

KURTOSIS OF $1_{yt}$ BEFORE AND AFTER DIVISION
BY FORECAST STANDARD DEVIATION $\hat{s}_t = \sqrt{\alpha + \beta A_t}$

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Entire Sample Period August 1873-December 1971</th>
<th>Reserved Subsample January 1951-December 1971</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kurtosis of $1_{yt}$</td>
<td>Kurtosis of $1_{yt}/\hat{s}_t$</td>
</tr>
<tr>
<td>1</td>
<td>1175</td>
<td>14.79</td>
</tr>
<tr>
<td>2</td>
<td>587</td>
<td>12.33</td>
</tr>
<tr>
<td>3</td>
<td>391</td>
<td>6.78</td>
</tr>
<tr>
<td>4</td>
<td>293</td>
<td>5.64</td>
</tr>
<tr>
<td>5</td>
<td>234</td>
<td>5.21</td>
</tr>
<tr>
<td>6</td>
<td>195</td>
<td>6.01</td>
</tr>
</tbody>
</table>
TABLE 4

KURTOSIS OF $1^y_t$ BEFORE AND AFTER DIVISION
BY THE PREDICTED STANDARD DEVIATION $1^s_t = \sqrt{\alpha + \beta \ell B_t}$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample Size</td>
<td>Kurtosis of $1^y_t$</td>
</tr>
<tr>
<td>1</td>
<td>1162</td>
<td>14.83</td>
</tr>
<tr>
<td>2</td>
<td>581</td>
<td>12.36</td>
</tr>
<tr>
<td>3</td>
<td>387</td>
<td>6.79</td>
</tr>
<tr>
<td>4</td>
<td>291</td>
<td>5.62</td>
</tr>
<tr>
<td>5</td>
<td>232</td>
<td>5.24</td>
</tr>
<tr>
<td>6</td>
<td>194</td>
<td>5.98</td>
</tr>
</tbody>
</table>

*aFor $\ell = 6$, where the estimate of $\alpha$ was negative, $\hat{\alpha}$ was set to zero.*
little true explanatory power that it is possible for them to provide. Thus, the poor reduction in kurtosis is not necessarily to be blamed on a change in the regression parameters from the regression sample to the reserved sample, but rather on the relative stability of variance, for whatever exogenous reason, in the latter period. This interpretation is confirmed by the fact that if the reserved sample is extended backwards by only three years, so as to cover the period 1947-1971 (as was done by mistake in one of the computations), fluctuation in variance is introduced into the sample, the sample kurtosis rises, and the lagged moving average does succeed in substantially reducing the kurtosis.

6. SUMMARY AND DISCUSSION

In Section 2, the kurtosis of a series of random variables was shown to increase in proportion to the square of the coefficient of variation of the variances of the individual random variables. In Section 3, many exogenous factors were suggested which can reasonably be expected to influence the variance of stock price changes. As a consequence, it was argued that the correct model to determine the variance of price changes may involve the effects of several exogenous factors, an autoregression on lagged values of the variance, and a distributed lag on the realized squared price changes. A procedure to estimate such a relationship by regression was explained. Since the appropriate explanatory variables to represent the exogenous factors have not yet been constructed, and since the lagged values of the variance are unobservable, the closest one can come to estimating the true model at the present time is to use lagged squared price changes as explanatory variables. A crude version of this approach, the use of a lagged moving average, provides a benchmark, since improved specification of the model can only increase the
explanatory power.

This device was applied to a series of monthly price changes in the New York Stock Exchange, as defined in Section 4. The results in Section 5 establish conclusively the existence of forecastable fluctuations in variance. Indeed, 70 percent of the deviation of the sample kurtosis from normality can be explained by the forecasted variances. The still better results for a two-sided moving average, where 86 percent of the deviation of the kurtosis from normality is explained, and the low values of the consistent upper bounds on the true kurtosis of the random variables (computed for the lagged moving average model, with and without an errors-in-variables adjustment) suggest that better forecasting models for the variance will explain virtually all of the nonnormality in the empirical frequency distribution of NYSE stock price changes.

These results suggest a new way of looking at the probability distribution of stock prices. Previously, the empirical frequency distribution observed over many periods was presumed to be a realization of the distribution prevailing at each instant in history. The long tails in this distribution were explained either by a nonnormal, infinite-variance stable distribution [11] or by the hypothesis of fluctuating variance in a normal distribution [14], [15]. The suggestion of fluctuating variance is not new; it appeared at least as early as 1961 in a remark by Working [2, p. 191]. What is new in the present results is the fact that the fluctuations in variance can be modeled and forecasted. Instead of the probability distribution at any one time having the same long tails as the observed empirical frequency distribution, as would be the case if the variance fluctuation were random and unpredictable, the price change actually has a nearly normal distribution at any
one time, with a variance which can be forecasted using an appropriate model. The apparent kurtosis of the empirical frequency distribution is the result of mixing distributions with predictably differing variances. Thus, the study of the probability distribution of security price changes must be broken into two parts: (i) modeling the fluctuations in variance in the series, and (ii) analyzing the distribution of price changes net of these predictable variance fluctuations. The outlines for this two-part approach are provided in this article, but much research in this area remains.

The results of the experiment have widespread implications for financial management and the theory of security markets. Some of these are the following: (i) the requirement for forecasts of price variance; (ii) the opening of the study of the determinants of price variance as a field of economic analysis; (iii) the need to respond to fluctuations in variance in portfolio management; (iv) the role of fluctuations in variance, through their effect on the riskiness of investment and, hence, on the appropriate risk premium, as an influence on the price level.

These innovations in finance will be important only to the degree that variance can be expected to exhibit forecastable and significant fluctuations in the future. In the NYSE data, very large forecastable fluctuations occurred in the first half of this century, but the variance of monthly price changes has been relatively stable thus far in the second half. It is important to anticipate the extent of future fluctuations in variance in the NYSE and other financial markets.