Research Program in Finance
WORKING PAPER SERIES

WORKING PAPER NO. 71

SYSTEMATIC RISK OF THE CRSP EQUAL-WEIGHTED COMMON STOCK INDEX: A HISTORY ESTIMATED BY STOCHASTIC-PARAMETER REGRESSION

by

James Ohlson
and
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.

Institute of Business and Economic Research
University of California Berkeley
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.
ABSTRACT

The systematic risk coefficient or "beta" of the equal-weighted index, with respect to the capitalization-weighted index, is studied. Stochastic-parameter regression serves to identify fluctuations in beta over a fifty-year history. Two kinds of stochastic variation in beta are found: a stationary first-order autoregressive process, which produces wide excursions about the mean value; and a serially independent random increment. Deviations from constancy are highly significant. Many features of the stochastic-parameter regression are illustrated in this application, most of which have not been used before.

I. Systematic Risk of the CRSP Equal-Weighted Index

I.1. Equal-Weighted and Capitalization-Weighted Indices

In any month, \( t \), the investment return on a common stock, \( n \), is the ratio of the end-of-month value of the investor's holding to the initial cost. Assuming that dividends are received at the month end, investment return is:

\[
1 + i_{nt} = \frac{p_{nt} + \text{div}_{nt}}{p_{n,t-1}}
\]

(1)

where \( 1 + i \) is total return, \( i \) is the "rate of return," \( p_{ns} \) is stock price at the end of month \( s \), and \( \text{div}_{nt} \) is dividends paid in month \( t \). (Prices and dividends must be computed on the basis of comparable shares

The authors are, respectively, associate professor and professor, School of Business Administration, University of California, Berkeley. This research was supported by NSF Grant GS3306. The valuable assistance of Jaffa Yehudai is gratefully acknowledged, as is Ellen McGibbon's help in preparing the manuscript.
ABSTRACT

The systematic risk coefficient or "beta" of the equal-weighted index, with respect to the capitalization-weighted index, is studied. Stochastic-parameter regression serves to identify fluctuations in beta over a fifty-year history. Two kinds of stochastic variation in beta are found: a stationary first-order autoregressive process, which produces wide excursions about the mean value; and a serially independent random increment. Deviations from constancy are highly significant. Many features of the stochastic-parameter regression are illustrated in this application, most of which have not been used before.

I. Systematic Risk of the CRSP Equal-Weighted Index

I.1. Equal-Weighted and Capitalization-Weighted Indices

In any month, \( t \), the investment return on a common stock, \( n \), is the ratio of the end-of-month value of the investor's holding to the initial cost. Assuming that dividends are received at the month end, investment return is:

\[
1 + i_{nt} = \frac{(p_{nt} + \text{div}_{nt})}{p_{n,t-1}},
\]

where \( 1 + i \) is total return, \( i \) is the "rate of return," \( p_{ns} \) is stock price at the end of month \( s \), and \( \text{div}_{nt} \) is dividends paid in month \( t \).

(Prices and dividends must be computed on the basis of comparable shares)

The authors are, respectively, associate professor and professor, School of Business Administration, University of California, Berkeley. This research was supported by NSF Grant GS3306. The valuable assistance of Jaffa Yehudai is gratefully acknowledged, as is Ellen McGibbon's help in preparing the manuscript.
and are adjusted for stock splits, stock dividends, or other capital transactions.

For any set of weights, \( w_{nt}, n = 1, \ldots, N \), that sum to one, an "index" of common returns can be defined as \( \sum_{n} w_{nt} (1+i_{nt}) \). The pivotal index, from the point of view of the Sharpe-Lintner-Mossin Capital Asset Pricing Model (see Jensen [1972]), is the "market" return, defined as the aggregate return to all assets. This index is computed by weighting each asset in proportion to the market value of shares outstanding, or "market capitalization" of the asset. Thus, if \( w_{Mn} \) is the market weight of each common stock, defined as \( w_{Mn} = \frac{\text{market value of outstanding common stock}}{\text{aggregate market value of all stocks}} \), then the market rate of return is defined by:

\[
i + i_{Mt} = \sum_{n=1}^{N} w_{Mn} (1+i_{nt}) .
\]

In principle, all assets should appear in the market portfolio. Limitations on data have slowed the development of a complete index. The broadest index for which an extensive history is available is the Center for Research in Security Prices (CRSP), University of Chicago, capitalization-weighted index of all common stocks listed on the New York Stock Exchange, available monthly from January 1926 to the present (Fisher and Lorie [1964]). In this index, "MW," each stock listed on the NYSE at the beginning of the month, is weighted in proportion to its market capitalization at the previous month-end.

Computation of the MW index requires data on the number of shares outstanding, as well as monthly stock returns. In the CRSP project, the
data on returns were collected first. With data on returns available, but without market weights, the natural index to compute was the equal-weighted "EW" index:

\[ \hat{\mu}_t = \frac{1}{N} \sum_{n=1}^{N} (1+i_{nt}) \]

The EW index is an imperfect substitute for MW. For one thing, it has certain logical inconsistencies: if two companies merge, the new entity, although comparable to the sum of the two previous companies in stature, is given one-half the weight; conversely, if a company is broken up, its components acquire added weight. The equal-weighted index stresses small companies, of which there are many more than large companies, to an extreme degree. For example, outstanding shares of the two largest American Companies, IBM and AT&T, have greater market value than the aggregate of the 800 smallest companies listed on the NYSE; nevertheless, the equal-weighted index gives 400 times as much weight to the latter companies. Hence, the EW index is not necessarily suited to serve as a proxy for the "market return" in studies of common stock returns.

Several important pioneering studies of capital asset pricing and investment performance (surveyed in Jensen [1972]) were conducted when EW was available, but MW had not yet been reconstructed. EW was used perforce as the "market return." These studies profoundly affected the literature. After MW became available, many researchers continued to use EW. In some cases, EW was retained to achieve comparability with earlier studies (e.g., Rosenberg and Marathe [1978]). In other cases, it may have been retained as a matter of inertia.
1.2. Systematic Risk

The central role of the market return is as the "systematic" component of common stock returns. For any asset, let the return in excess of the risk-free rate of return, or "excess return" be:

\[ r_{nt} = (1 + i_{nt}) - (1 + i_{Ft}) = i_{nt} - i_{Ft} \]

where \( i_{Ft} \) is the rate of return on a riskless asset (taken as the annualized return on ninety-day Treasury bills in this study). Similarly, the excess market return is \( r_{Mt} = i_{Mt} - i_{Ft} \). Then, the excess return on any asset may be expressed as:

\[ r_{nt} = \alpha_{nt} + \beta_{nt} r_{Mt} + u_{nt}, \]

where \( \beta_{nt} \) is the "systematic risk coefficient," commonly known as "beta"; \( \alpha_{nt} \) is the "abnormal return," which is zero under the simplest version of the Capital Asset Pricing Model; and \( u_{nt} \) is the "residual return," uncorrelated with the systematic component of return (\( r_{Mt} \)).

Alternatively, the constructs \( r_{nt} \) and \( r_{Mt} \) may be defined as:

\[ (4*) \quad r_{nt} = \ln(1+i_{nt}) - \ln(1+i_{Ft}) \quad ; \quad r_{Mt} = \ln(1+i_{Mt}) - \ln(1+i_{Ft}) \]

This logarithmic form of equation (4) has advantages and disadvantages, which are discussed in Rosenberg and Marathe [1978, section II.A.3].

For present purposes, the crucial advantage is that residual returns in the logarithmic model more closely obey a normal distribution. Since maximum likelihood estimation and likelihood ratio tests presuming
this distribution are used below, the logarithmic definition of excess returns (4*) is used henceforth.

The equal-weighted logarithmic excess return is defined analogously:

\[ r_{Et} = \ln(1+i_{Et}) - \ln(1+i_{Ft}) . \]

If (5) is correctly specified, with \( r_{Mt} \) as the explanatory variable, then use of \( r_{Et} \) is a misspecification. The extent of bias due to the erroneous explanatory variable depends on the joint distribution of \( r_{Et} \) and \( r_{Mt} \). Following the CAPM, the expected relationship between EW and MW is of the form:

\[ r_{Et} = \alpha(t) + \beta(t)r_{Mt} + \nu_t , \]

where \( \alpha \) and \( \beta \) are written as functions of time, since they may not be constant. If \( \alpha(t) \) is identically zero, \( \beta(t) \) is constant, and \( \nu_t \) is identically zero, then \( r_{E} \) and \( r_{M} \) are exactly proportional and \( r_{E} \) is a perfect proxy: substitution of \( r_{E} \) for \( r_{M} \) then leaves the properties of (5) unchanged except for a change of scale in beta. On the other hand, variance of \( \alpha(t) \) and \( \nu_t \) or nonstationarity of \( \beta(t) \) cause "errors in variables" in (5), with resulting downward estimation bias in the slope coefficient (Miller and Scholes [1972]) and upward bias in the intercept \( \alpha \). A nonzero mean for \( \alpha(t) \) also biases the intercept in (5). The properties of (7) are therefore guides to the impact and severity of the misspecification problem in (5).
The behavior of the equal-weighted index is also of intrinsic interest. A number of professional investors have considered an equal-weighted portfolio to be an attractive investment policy, because it has historically outperformed the capitalization-weighted market return. The first "index fund" known to us (an investment portfolio designed to reproduce the performance of an index) was designed to match an equal-weighted index (see Black and Scholes [1974]). The parameters of (7) define the systematic risk coefficient, $\beta(t)$, abnormal performance, $\alpha(t)$, and residual risk, $\text{VAR}(v_t)$, of an equal-weighted strategy. Hence, they bear directly on the desirability of such a strategy.

1.3. A Stochastic Model

Logarithmic excess returns for the 600 months from January 1926 to December 1975 will be analyzed. The parameters $\alpha(t)$ and $\beta(t)$ are of primary interest. Ideally, one would identify causal variables or concomitants that would explain changes in these parameters over time. Lacking such variables, the present study employs stochastic-parameter regression to estimate the history of parameter values. The stochastic-parameter model should be chosen so as to accord nonnegligible probability to the actual historical pattern of the parameters, so that maximum likelihood estimation can successfully uncover this pattern, or one like it, as the most probable. To frame an appropriate stochastic model, it is useful to survey possible causes of parameter change and the likely results of these.

Consider, first, the systematic risk coefficient, beta. To bring economic theory to bear on the question, suppose that there are a number
of economic events or "common factors" that may influence the returns on many companies. Every company exhibits an intrinsic response to each factor. The weighted average of these responses is the index response to each factor. Then, for each factor, the relative response coefficient equals the ratio of the responses of EW and MW indexes. The systematic risk of EW is a weighted average of the relative response coefficients, weighted in proportion to the variances of the associated factors (Rosenberg [1974]). The greater the importance (variance) of a factor as a component of return, the more important is relative response to that factor as a determinant of beta (Rosenberg and Gru [1976]).

Changes in ex ante variance of common factors therefore change the weights of the relative response coefficients, and hence change beta. Determinants of these changes would be understood in the context of a macroeconomic model, beyond the scope of this study.

Change in any relative factor response coefficient also affects beta. Changes in the composition of weights of EW and MW can cause this. Major change occurs when weights of a group of high-response or low-response companies increase in one index relative to the other. Additions to and deletions from the NYSE list are one source of changing weights. Also, mergers and breakups influence weights in EW, while market price changes influence weights in MW. Changes in concentration within an industry alter the weight of the industry in EW, while changes in industry importance or profitability, reflected in changed market value, alter the weight in MW.
Without going into further detail, one characteristic that emerges clearly is that beta changes are likely to exhibit serial dependence. In other words, a change—due to any of the above causes—that occurs in a given month is likely to persist over several months and, indeed, may persist over years or decades. Thus, the stochastic model should exhibit "memory," so that the value arrived at in the previous period influences the current value.

There is a plausible argument leading to the conclusion that the process may be stationary: The decision to list a stock on the NYSE is a judgmental one. The distribution of companies by size is controllable through merger and divestiture. The response of common stock return to business risk in a given company is controllable by adjusting leverage in the financial structure. Thus, through modifications in rather inconsequential characteristics of business enterprise, the distribution of common stocks in regard to systematic risk can be controlled. It follows that if investor preferences suggest a certain optimum distribution, there are relatively small economic costs to reorganization of firms to satisfy that preference. Thus, it would not be surprising to find a tendency for the systematic risk of EW to regress gradually to some normal value after stochastic shocks had caused it to deviate therefrom. Mechanisms for such adjustments might be the NYSE listing process; the net creation of smaller companies through birth, death, or merger; or changes in financial leverage of small companies that adjust common stock investment risk relative to the purely "business risk" intrinsic in their unlevered earnings.
A continuous model with memory that admits regression toward the norm as a special case is the second-order autoregressive process, with parameters \( \phi \) and \( \phi' \):

\[
\beta_t = \beta_0 + \phi \beta_{t-1} + \phi' \beta_{t-2} + d_t,
\]

where \( d_t \) is a random shift which is serially independent, with mean zero in all periods and variance \( \sigma^2 \). The model may be rewritten, provided that \( \phi + \phi' \neq 1 \), in the form:

\[
\beta_t = \bar{\beta} + \phi(\beta_{t-1} - \bar{\beta}) + \phi'(\beta_{t-2} - \bar{\beta}) + d_t,
\]

where \( \bar{\beta} = \beta_0 / (1 - \phi - \phi') \). In this form, the model is seen to imply the mean value \( \bar{\beta} \), with deviations \( \delta_t = (\beta_t - \bar{\beta}) \) obeying a second-order autoregressive process:

\[
\delta_t = \phi \delta_{t-1} + \phi' \delta_{t-2} + d_t.
\]

As another special case, the model produces a Martingale (when \( \beta_0 = \phi' = 0 \) and \( \phi = 1 \)):

\[
\beta_t = \beta_{t-1} + d_t.
\]

The tendency for drift in beta does not exhaust the possible forms of stochastic variation. The response parameters of EW and MW may also vary from period to period in a random fashion, due to the unique nature of events in each period. Each period can be viewed as something in the nature of a random drawing from a distribution of possible events, each
with different factor responses implying different systematic risk for EW. Hence, beta in each period may exhibit a purely random increment, $\varepsilon$, with mean zero and stationary variance $\Lambda$. The model including this phenomenon may be written:

$$ r_{Et} = \alpha + (\beta_t + \varepsilon_t) r_{Mt} + v_t. $$

The contribution of $\varepsilon$ may be merged with the disturbance into a composite return:

$$ r_{Et} = \alpha + \beta_t r_{Mt} + u_t, \quad u_t = \varepsilon_t r_{Mt} + v_t, $$

where:

$$ \mathbb{E}[u_t] = 0; \quad \text{VAR}[u_t | r_{Mt}] = \sigma^2 + r_{Mt}^2 \Lambda; \quad \text{COV}[u_t, u_s] = 0 \text{ for } t \neq s. $$

This formulation clarifies the role of the random increment to beta: its effect is indistinguishable from a variance component that is proportional to squared market return.

The stochastic process for alpha is less clearly suggested by economic theory. Capital market theory suggests that the presence of a significant nonzero alpha at any time creates the opportunity for speculative profit. Thus, a nonzero alpha is self-disfulfilling and presumably arises from disequilibrium in capital markets. Extensions to the Capital Asset Pricing Model (Jensen [1972]), which admit more realistic assumptions on borrowing lending, and taxation, do lead to small nonzero alphas. In particular, when the available zero-beta return differs from the riskless rate posited in the model, the misspecification results in an alpha
that is a linear function of beta. Therefore, lacking any other guidance, it is natural to admit a stochastic process for alpha that is identical to that for beta. However, a random increment to alpha is observationally equivalent to residual return $v_t$, and hence unidentifiable. Thus, a second-order autoregressive process for alpha of form (8) is all that is needed.

1.4. Preliminary Simplification

Before analyzing the equal-weighted index, we applied the methodology to the history of returns on an individual common stock. The complete model includes seven "specification parameters": $\phi$, $\phi'$, and $Q$ for the beta process; three analogous parameters for the alpha process; and the variance, $\Lambda$, of the random increment to beta. We found that the data were uninformative concerning the second-order coefficient $\phi'$ in the beta process and the three parameters of the alpha process. The sample likelihood is essentially unchanged when these parameters are varied over the entire reasonable range. The difficulty occurs because residual variance, $\text{VAR}(u_t)$, is very large: for a typical asset, residual standard deviation is roughly .08 per month. A substantial value for alpha, say .005 per month, completely disappears in the background noise, unless it persists over a very long interval. Hence, a stochastic process for alpha is completely obscured. A sizable change in beta, say .20, contributes a return with magnitude .016 in a typical month (since the root mean square market return is about .08). This contribution is larger in comparison to residual return but does not emerge.
distinctly. Since the systematic returns in any single month or series of consecutive months are not clearly distinguished from residual return, the added flexibility of specification provided by a second-order beta process cannot greatly increase the probability of the fitted beta history.

We anticipated that data for the equal-weighted index would also be essentially uninformative concerning these parameters. Index residual variance is only one-third as great but is still sufficient to obscure these aspects of the parameter processes. We therefore chose to fit a simplified model that omits the poorly distinguished parameters:

\begin{equation}
    r_{Et} = \alpha + \bar{\beta} r_{Mt} + \delta_t r_{Mt} + u_t \quad t = 1, \ldots, 600
\end{equation}

\begin{equation}
    \delta_t = \phi \delta_{t-1} + d_t \quad t = 2, \ldots, 600
\end{equation}

\[ E[u_t] = 0 \; ; \; \text{VAR}[u_t] = \sigma^2 + \gamma^2 r_{Mt}^2 \Lambda \; ; \; E[d_t] = 0 \; ; \; \text{VAR}[d_t] = Q. \]

The model has three "specification parameters": \( \phi \), \( Q \), and \( \Omega \). When \( |\phi| < 1 \), the beta process is stationary, \( \bar{\beta} \) is the expected value for \( \beta_t \), and the stationary variance of \( \delta_t \) is:\footnote{The stationary variance for \( \delta \) is found by observing that \( \text{VAR}(\delta_t) = \phi^2 \text{VAR}(\delta_{t-1}) + \text{VAR}(d_t) \) and solving for \( \Omega = \text{VAR}(\delta_t) = \text{VAR}(\delta_{t-1}) \).}

\begin{equation}
    \text{VAR}(\delta_t) \equiv \Omega = Q/(1 - \phi^2) .
\end{equation}

The initial condition for the beta process is the prior distribution for beta in period 1. Since \( \bar{\beta} \) is a fixed but unknown parameter, but \( \delta \) is
a stationary process, initialization consists in applying the stationary distribution as the prior distribution for $\delta_1$:

$$E[\delta_1] = 0 \quad ; \quad \text{VAR}[\delta_1] = \Omega.$$  

If, on the other hand, $\phi = 1$, then the process is a Martingale, and the distribution for $\beta$ is nonstationary. Then the fixed but unknown parameter, $\bar{\beta}$, may be used as the starting value for beta in period 1, with $\delta_1 = 0$.

The processes $u$ and $d$ are assumed to be normally distributed and to be uncorrelated with one another and independent of the market return.

II. Estimating the Stochastic Specification

II.1. Methodology

The model (13) has six parameters: $\phi$, $Q$, $A$, $\sigma^2$, $\alpha$, and $\bar{\beta}$. For purposes of estimation, it is useful to divide these into two groups: three "specification parameters," $\phi$, $Q$, and $A$; and the three usual regression parameters, $\sigma^2$, $\alpha$, and $\bar{\beta}$. Four different versions of the specification parameters are used as indicated by the transforms in table 1.

In the transformation to $\Theta$, the variances $Q$ and $A$ are expressed in relative terms, as multiples of the disturbance variance $\sigma^2$. This causes $\sigma^2$ to become the scale parameter for all second moments in the model, so that it enters linearly in the variance matrix of the
<table>
<thead>
<tr>
<th>Natural</th>
<th>Relative \ Variances</th>
<th>Unrestricted</th>
<th>Transformed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi$</td>
<td>$\phi = \theta_1$</td>
<td>$\phi = z_1^2/(1 + z_1^2)$</td>
<td>$\phi$</td>
</tr>
<tr>
<td>$Q$</td>
<td>$Q = \sigma^2 \theta_2$</td>
<td>$Q = \sigma^2 z_2^2$</td>
<td>$\rho = \phi \Omega/(\Omega + \Lambda)$</td>
</tr>
<tr>
<td>$\Lambda$</td>
<td>$\Lambda = \sigma^2 \theta_3$</td>
<td>$\Lambda = \sigma^2 z_3^2$</td>
<td>$\Lambda = \sigma^2 \Lambda^*$</td>
</tr>
</tbody>
</table>

data, which may be written as $\sigma^2 \sim (\theta)$. The purpose of this is to reduce from four to three the dimensionality of the space for parameter search by numerical procedures. The log likelihood function may be condensed as a function of $\theta$ only, $\ell(\theta)$, such that $\ell$ is easily computed for given $\theta$. \(^1\)

The parameter vector, $z = (z_1 : z_2 : z_3)'$, is used to define the specification during maximum likelihood estimation. As the vector $z$ ranges over three-dimension Euclidean space, the specification varies over the admissible region: $0 \leq \phi < 1$ ; $0 \leq Q$ ; $0 \leq \Lambda$. Hence,

\(^1\) In other words, the maximum of the log-likelihood function is obtained by a two-step procedure:

$$\max \ell = \max \{ \max_{\theta, \sigma^2, \alpha, \beta} \ell(\theta, \sigma^2, \alpha, \beta) \} \equiv \max \ell(\theta),$$

where the first (conditional) maximization is solved by a GLS procedure (see below), and the second maximization is solved by a nonlinear optimization program.
unconstrained maximization over $R(z)$ corresponds to a constrained maximization over $R(\theta)$. Thus, by use of $z$, optimization programs for unconstrained problems may be used.\footnote{The transformation from $z$ to $\theta$ is many-one, and it is symmetrical about zero. This causes spurious local extrema of the log likelihood at $z=0$. A simple adjustment to the optimization program might be needed in some cases to avoid cycling near these extrema, but none was required in our applications.}

Finally, in the "transformed" parameter vector, $Q$ is replaced by the first-order autocorrelation coefficient, $\rho$, of total systematic risk. The coefficient is defined in terms of $\Omega$, the stationary variance of the sequential beta process, and $\Lambda$, the variance of the random process, by\footnote{The derivation of $\rho$ is thus as follows:}

\begin{equation}
\rho = \text{corr}(\bar{\beta} + \delta_t + \varepsilon_t, \bar{\beta} + \delta_{t-1} + \varepsilon_{t-1}) = \frac{\phi\Omega}{\Omega + \Lambda}.
\end{equation}

Also, the variance of the random beta increment, $\Lambda$, is expressed relative to disturbance variance by $\Lambda^*$. The purpose of the transformation is to obtain specification parameters with respect to which the log likelihood function approximates a quadratic form. This facilitates more rapid convergence of an optimization program (a feature that was not
exploited in this study) and simplifies the preparation of confidence regions (see section II.3, below).

The estimation procedure is based on the decomposition of $\mathcal{L}$. For given $\theta$, the stochastic parameter regression problem may be shown to be equivalent to a generalized least squares (GLS) regression, with the variance matrix for the data, $V(\theta)$, dependent upon $\theta$. The problem, then, is to evaluate the familiar likelihood function for a GLS regression, which involves computing the inverse of $V$ (a $600 \times 600$ matrix in this problem). The computational problem is solved by exploiting the recursive structure of the stochastic parameter process. The process is naturally expressed as a sequence of prior distributions for successive periods, each based on the previous period's parameter values. As explained in Rosenberg [1973], the formulation results in a recursive expansion of $\mathcal{L}$, which diagonalizes the variance matrix. Thus, $\mathcal{L}$ (equation (29) in Rosenberg [1973]) is evaluated by a set of recursive formulas (simplifications of equations (37a-37p) in Rosenberg [1973]), which are processed beginning with period 1 and proceeding period by period. The computational burden is bearable: each evaluation requires roughly three times as many computations as an ordinary least squares regression. By contrast, direct inversion of $V$ would require approximately 36,000 times as many computations.

The estimation procedure consists of a search of the admissible region to find the maximum likelihood specification, as well as to obtain likelihood values at other interesting specifications. We employed an algorithm of Brent [1972, Ch. 7], which does not use analytic partial
derivatives. Convergence was quite satisfactory. The admissible region was searched widely to verify that there were no local extrema.

II.2. Estimates of the Specification for Beta

Table 2 reports maximum likelihood estimates (MLEs) for six possible model specifications. Model 1 is the complete model (equation (13)), with an unrestricted first-order process, plus an additive random increment. The MLE for the convergence rate, \( \phi \), is a very slow .9875 per month: the half-life of each random shift \( d_t \) (the time required before only one-half of that shift survives) is therefore fifty-five months. The variance \( Q \) of the shifts is .000655, while the variance \( A \) of the random increment \( \epsilon_t \) is quite large: .0839.

Model 2 is restricted to \( \phi = 1.0 \) and therefore implies a Martingale for \( \beta_t \), plus the additive random increment. Column 4 shows that maximum log likelihood for this model falls 5.67 short of that achieved for model 1. According to the asymptotic distribution for the likelihood ratio, since the maintained hypothesis that \( \phi = 1.0 \) restricts one parameter of the model, twice this difference is distributed as \( \chi^2 \) with one degree of freedom. The critical value at the 99.5 percent level is 7.88. \( 2(\hat{\ell}_{\text{max}} - \ell) = 11.38 \) exceeds the critical value, so the Martingale hypothesis may be rejected at the 99.5 percent level of confidence.

Models 3 and 4 involve other restricted values for the convergence rate: \( \phi = 0.8 \) and \( \phi = 0.5 \), respectively. Model 5 sets \( \phi = 0.0 \), so that the first-order process becomes a process without memory, indistinguishable from the random serially independent increments \( \epsilon_t \) also present.
<table>
<thead>
<tr>
<th>Model for Beta</th>
<th>Convergence Rate</th>
<th>Variance of Shift</th>
<th>Random Variance</th>
<th>Deficiency, in Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sequential + random variation</td>
<td>.9875</td>
<td>.000665</td>
<td>.0839</td>
<td>0</td>
</tr>
<tr>
<td>2. Martingale (ϕ = 1.0) + random</td>
<td>set at 1.0</td>
<td>.000420</td>
<td>.0847</td>
<td>5.67</td>
</tr>
<tr>
<td>3. Sequential (ϕ = 0.8) + random</td>
<td>set at 0.8</td>
<td>.0149</td>
<td>.0668</td>
<td>6.21</td>
</tr>
<tr>
<td>4. Sequential (ϕ = 0.5) + random</td>
<td>set at 0.5</td>
<td>.0520</td>
<td>.0364</td>
<td>8.44</td>
</tr>
<tr>
<td>5. Random variation only</td>
<td>set at 0.</td>
<td>set at 0.</td>
<td>.1030</td>
<td>13.59</td>
</tr>
<tr>
<td>6. Constant</td>
<td>set at 0.</td>
<td>set at 0.</td>
<td>set at 0.</td>
<td>102.87</td>
</tr>
</tbody>
</table>
in the model. To avoid the identification problem, one or the other of the two processes must be removed. This is done by setting the shift variance \( \Omega \) to zero also in model 5. Thus, model 5 admits only random variation in beta. As \( \phi \) declines toward zero, \( \lambda \) falls, and the models are more and more strongly rejected.

The final model eliminates random variation as well and imposes a constant beta. That is, \( \gamma = \Omega \), and the model specifies a simple linear regression. There are three restrictions on \( \theta \), so twice the deficiency in log likelihood is asymptotically distributed as \( \chi^2 \) with three degrees of freedom. The critical value at the 99.5 percent level of confidence is 12.84; twice the log likelihood deficiency is 205.74, so beta constancy is overwhelmingly rejected.

Table 3 shows contributions to the variance of beta and \( r_E \), as implied by the six models. For models 1, 3, and 4, the first-order process is stationary, with variance \( \Omega \) given in (14). The Martingale model is nonstationary and no variance is defined. Estimated \( \lambda \) increases as \( \Omega \) decreases, so that implied total variance for beta is nearly constant. When this fact is taken in conjunction with the likelihoods of the different models, a clear picture emerges: the most crucial improvement in goodness of fit is obtained by allowing for variation in beta, whether random or sequential. The best fit is obtained with a variance estimate of roughly .11 (corresponding to a standard deviation of .33). This accounts for most of the large increase in \( \lambda \) from model 6 to any of models 1 through 5. Further improvement in fit results from establishing the autocorrelation function
for beta as reflecting the sum of a purely random process and an autoregressive process with very long half-life.

**TABLE 3**

COMPONENTS OF BETA VARIANCE AND VARIANCE OF $r_E$

<table>
<thead>
<tr>
<th>Model</th>
<th>Variance of Beta</th>
<th>Variance of $r_E$</th>
<th>$V$</th>
<th>$\Delta$</th>
<th>$\Omega$</th>
<th>$\Delta \Omega$</th>
<th>$\Delta \Omega^2$</th>
<th>$\Delta \Omega^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0267</td>
<td>0.0839</td>
<td>0.1106</td>
<td>0.004838</td>
<td>0.000093</td>
<td>0.000294</td>
<td>0.000311</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>nonstationary</td>
<td>0.0847</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.000297</td>
<td>0.000311</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0415</td>
<td>0.0668</td>
<td>0.1083</td>
<td>0.005036</td>
<td>0.000145</td>
<td>0.000234</td>
<td>0.000312</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.0694</td>
<td>0.0364</td>
<td>0.1058</td>
<td>0.005089</td>
<td>0.000243</td>
<td>0.000128</td>
<td>0.000316</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0.1030</td>
<td>0.1030</td>
<td>0.005143</td>
<td>0</td>
<td>0.000361</td>
<td>0.000322</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.005542</td>
<td>0</td>
<td>0</td>
<td>0.000689</td>
<td></td>
</tr>
</tbody>
</table>

The latter process accounts for only a fraction of total beta variance ($0.0267/0.1106 = 0.241$), but it is nevertheless of great interest for two reasons: first, it is forecastable; second, it is persistent over time. In actual investment strategy, frequent transactions and beta adjustments are rare, so that a prediction of average beta over a long horizon is relevant. The sequential process, with its long half-life, persists with little attenuation, while the random variation averages to zero over a longer interval. The sequential component amounts to only 24 percent of beta variance in a single month, but accounts for more than 75 percent of interyear variance in average annual beta.
It is also interesting that the first-order autocorrelation coefficient \( \rho \) for beta is relatively constant across three models with stationary first-order processes: .24, .31, and .33 for models 1, 3, and 4, respectively. This occurs because the decline in \( \phi \) is nearly offset by the increase in \( Q \).

The difference between the equal-weighted index return and its mean may be written as follows:

\[
\begin{align*}
\tau_{E_t} - E[\tau_{E_t}] &= \alpha + (\beta + \delta_t + \varepsilon_t)\tau_{M_t} + u_t - E[\alpha + (\beta + \delta_t + \varepsilon_t)\tau_{M_t} + u_t] \\
&= \beta (\tau_{M_t} - E(\tau_{M_t})) + (\delta_t + \varepsilon_t)\tau_{M_t} + u_t.
\end{align*}
\]

Since \( \alpha \) and \( \beta \) are constants, and \( E(\delta_t) = E(\varepsilon_t) = E(u_t) = 0 \), and the stochastic elements in \( \beta \) are uncorrelated with the market return, this becomes:

\[
(17) \quad \tau_{E} = \beta (\tau_{M} - E(\tau_{M})) + (\delta_t + \varepsilon_t)\tau_{M} + u_t.
\]

Hence, the variance of \( \tau_{E} \) is:

\[
\text{VAR}(\tau_{E}) = \beta^2 \text{VAR}(\tau_{M}) + (\text{VAR}(\delta_t) + \text{VAR}(\varepsilon_t))E(\tau_{M}^2) + \text{VAR}(u_t)
\]

\[
(18) \quad \text{VAR}(\tau_{E}) = \beta^2 \text{VAR}(\tau_{M}) + (\Omega + \Lambda)E(\tau_{M}^2) + \sigma^2.
\]

When \( \hat{\beta} \) (table 4, below) is substituted into (18), the decomposition of sample variance of \( \tau_{E} \) in the second panel of table 3 results. Total/variance of \( \tau_{E} \) is .00623. The interpretation of the results for model 1 is that 77.7 percent of the variance of \( \tau_{E} \) is attributable to the normal value of systematic risk, 1.49 percent to stationary sequential variation in beta, 4.71 percent to random fluctuation in beta, and 4.99 percent to
residual variance. The sum of the percentages is not 100. The balance, 11 percent of total variance, is due to historical interaction between beta and market variance. The expectation for this is zero, but, in the sample history below, beta is seen to be much higher during the depression years when market variance is larger.

II.3. The Shape of the Likelihood Function or Bayesian Posterior

The convergence rate $\phi$ is of special interest, since the model ranges from purely random beta through a stationary autoregressive process to a Martingale as $\phi$ ranges from 0 to 1. Figure 1 depicts twice the log likelihood throughout the admissible region. The curve is obtained by interpolation through a grid of values, which is quite dense in the region of the maximum. An asymptotic confidence region for $\phi$ is defined by the range within which $\ell$ falls short of $\ell_{\text{max}}$ by less than the critical value. The 99.5 percent confidence region is shown on the figure: it is the interval within which $2(\ell_{\text{max}} - \ell(\phi)) \leq \chi^2_{.005} = 7.87$. The region extends from $\phi = .905$ to $\phi = .995$. The half-life of random shifts varies from 6 months at the lower end of the region, through 55 months at the maximum likelihood estimate, to 137 months at the upper end.

From a Bayesian standpoint, with an uninformative prior distribution, the kernel of the posterior distribution for the specification parameters is closely related to the likelihood function. In fact, the two differ only by a correction for the determinant of the subspace spanned by the regressors (Rosenberg [1973], eq. (33)).
FIGURE 1

99.5% confidence region:
0.905 ≤ φ ≤ 0.9995
In this problem, the correction factor is relatively insignificant. The conditional posterior density for \( \phi \), given the MLE for \( Q \) and \( \Lambda \), is shown in figure 2. The mode of the Bayesian posterior effectively coincides with the MLE.

From table 1 it is apparent that, as \( \phi \) decreases, the constrained maximum is obtained at larger values of \( Q \) and smaller values of \( \Lambda \). To underscore this point, a pair of conditional densities for \( Q \) and another pair for \( \Omega \) are shown in figures 3 and 4. For \( Q \), the posterior densities are plotted for two values of \( \phi \) that are relatively close to one another: the MLE \( \hat{\phi} = .9875 \) and \( \phi = .999 \). (In each case, the chosen value of \( \Lambda \) is the MLE, given \( \phi \).) Although \( \phi \) changes little, the half-life increases from 55 months to more than 600. Hence, it is not surprising that both the posterior mode and mean of \( Q \) decrease.

For \( \Lambda \), posterior densities are shown for two extremely different models: first, model 1 with MLE for \( \phi \) and \( Q \); second, model 5 with \( \phi = Q = 0 \). The densities are slightly skewed but more nearly symmetric than the conditional densities for \( \phi \) and \( Q \).

It is common practice to posit that the log likelihood function (or kernel of the Bayesian posterior density) is quadratic in the region of the maximum, so that the matrix of second partials (Hessian) may be used as a basis for the variance-covariance matrix of estimation error. In the preliminary studies, we were concerned that the asymmetry of the conditional densities and the curving ridge relating \( \phi \) to \( Q \) and \( \Lambda \) would render the quadratic approximation invalid. To check this, we fitted a quadratic form in \( \phi \), \( Q \), and \( \Lambda \), with extremum

---

1 This was verified computationally for some specific cases.
FIGURE 2

A CONDITIONAL POSTERIOR DISTRIBUTION FOR $\phi$

$P(\phi | Q = .000665, \Lambda = .0839)$
FIGURE 3

TWO CONDITIONAL POSTERIOR DISTRIBUTIONS FOR $Q$

1. $P(Q|\phi = 0.9875, \Lambda = .0839)$
2. $P(Q|\phi = 0.999, \Lambda = .0846)$
FIGURE 4
TWO CONDITIONAL POSTERIOR DISTRIBUTIONS FOR $A$:
1. $P(A | \phi = 0.9875, Q = 0.000665)$
2. $P(A | \phi = 0, Q = 0)$
at the MLE, to the evaluations of the log likelihood function. The accuracy of the approximation is most important in the region of the maximum, since this is where the greatest probability mass lies. To reflect this, we fitted the quadratic only to evaluations of the likelihood function that fell within the 90 percent asymptotic confidence region for the parameters. Two hundred and twenty-five such evaluations were obtained. The least squares fit achieved an $R^2$ of only .760, confirming that the quadratic approximation was poor indeed.

We then considered whether a transformed parameter vector might allow improved fit. Only one transformation was tried: replacement of $Q$ by $\hat{\rho}$, and $A$ by $A^*$, as explained in table 1. This transformation was highly effective, for the $R^2$ improved to .928. Clearly, if the quadratic approximation is to be used to define estimation error variance and confidence ellipsoids, care must be taken to define the parameters appropriately. Other transformations might work still better. For example, $(\Omega + A)$ might replace $A^*$.

III. Regression Parameter Estimates

III.1. Estimates of the Unknown Parameters

Each model specification $\theta$ implies a variance matrix $V(\hat{\theta})$ for the observations, and a generalized least squares regression, in which the regression coefficients are $\alpha$ and $\bar{\beta}$, and the "scale parameter" for the variances and covariances of disturbances is $\sigma^2$. Using $\hat{\sigma}^2$ and $V(\hat{\theta})$, the standard errors of the coefficients may be computed.\footnote{From a Bayesian standpoint, it would be preferable to compute the conditional distribution for these parameters for each $\theta$ and then}
Table 4 contains the estimated regression parameters and standard errors for the three most important models: model 1, with sequential and random variation in beta; model 5, with random variation only; and model 6, assuming constant beta.

| Model | $\hat{\alpha}$ | $\hat{\beta}$ | $\hat{\beta}_{601|600}$ |
|-------|----------------|----------------|--------------------------|
| 1     | .00026         | 1.187          | 1.175                    |
|       | (.00087)       | (.080)         | (.104)                   |
| 5     | -.00010        | 1.223          | 1.223                    |
|       | (.00088)       | (.024)         | (.024)                   |
| 6     | .00058         | 1.270          | 1.270                    |
|       | (.00108)       | (.018)         | (.018)                   |

For each model, the standard errors are computed under the presumption that the specification is correct. Hence, if the model is misspecified, bias in the standard errors results. Comparing the three models, the standard error of alpha is seen to change little. This is not surprising. The constant term is almost orthogonal to the market return. When two dimensions are orthogonal, change of the

integrate with respect to the posterior density of $\theta$ to obtain the sampling distribution of the coefficients (Rosenberg [1973], eq. (23)). To escape this laborious numerical integration, we have instead computed the sampling theory for the parameters, under the presumption that the MLE specification is correct.
variance matrix along one dimension is not likely to affect the computed standard error in the other.

The estimate for alpha is less than the standard error in all cases. However, the differences in estimated alpha are nontrivial. The difference between model 1—which is preferred in this study—and model 6—which has been suggested in prior literature—is .00032. This refers to the logarithm of monthly return and therefore corresponds to .00384 in the logarithm of annual return, or a difference of approximately four-tenths of 1 percent in the annual rate of return (forty basis points). Compounded over the fifty-year data history, this becomes a difference of more than 20 percent in cumulative value of the investment! As in all cases of performance analysis in the stock market, the intrinsic noisiness of residual returns obscures the relative performance of an investment strategy, to the extent that important differences in performance are practically indistinguishable.

Interestingly, model 5 leads to a negative alpha. Acceptance of this model would lead to the conclusion that the equal-weighted strategy is inferior. With models 1 or 6, the performance appears to have been superior, but not statistically significant. Moreover, the estimated alpha is of the order of magnitude that could be expected from upward bias due to measurement error in prices (L. Fisher [1966], pp. 197-199). Thus, the results do not strongly recommend the equal-weighted strategy.

Turning to beta, the model predictions differ significantly. Where beta is assumed constant, it is estimated at 1.270 with ostensible
precision: the estimated standard error is only .018. When random variation is admitted, the estimate of mean beta falls to 1.223, and the standard error rises. When sequential variation is also admitted, the estimate falls to 1.187, with standard error of .080. Thus, mistaken adherence to the constant beta model would result in a misleading estimate for beta and an entirely wrong standard error. This conclusion—that OLS is misleading, with estimation accuracy severely overstated—also applied to our preliminary analysis of a single security.

The last column in table 4 compares predictions for beta in period 601. The subscript $T+1|T$ denotes a prediction based upon information through period $T$, with $T=600$ in this case. These predictions are based upon the best estimate for the parameter process in period $T$, $\hat{\beta}_T|T$, often called the "filtered estimate" (Rosenberg [1973], eq. (36)) extrapolated forward according to equations (19) to (22), below. Notice the substantial differences in forecasts. The increased standard error in model 1 stems from imperfect prediction of $\delta_{T+1}$.

III.3. The "Smoothed" History of Systematic Risk

For each model specification, there are maximum likelihood estimators of the realization of the stochastic process that underlies the data history. Let $\hat{\beta}_s|T$ be the estimator for any period $s$, $\hat{\beta}_s|T = \hat{\beta} + \delta_s|T$, subscripted $s|T$ to indicate that the entire history, through period $T$, is employed in the estimate. These are often called "smoothed" estimators, because of the smoothness of the time history.
The estimates are computed by recursive formulas (4-17) given in Cooley, Rosenberg, and Wall [1977].

Figure 5 shows the history. It begins with beta at 1.08 in the years before the Great Crash; then, beginning in 1929, there is a rapid increase, with a single small reversal, peaking at 1.41 in 1936. Beta remains high through the remaining depression years, declines in the early war years, increases briefly in 1942, and then begins a long, regular decline from 1.39 in 1943 to .92 in 1956. Then a steady rise carries beta to 1.28 in 1973, a level from which it has since declined to the December 1976 estimate of 1.175.

This figure also shows confidence bands for beta, drawn at

\[ \hat{\beta}_s \pm \sqrt{E(\hat{\beta}_s - \beta_s)^2} \].

The band width at each point in time constitutes a 68.3 percent confidence region.\(^1\)

The presence of sequential variation in beta has been confirmed by significant improvement in the likelihood function. It is also interesting to test for significant shifts in beta over particular intervals that correspond to important historical events. Prior to inspecting

\(^1\)The sampling theory is again approximated via the assumption that the MLE specification \(\hat{\theta}\) is correct. Under this assumption, it remains to estimate the scale parameter \(\sigma^2\) of second moments of stochastic terms. This raises an additional problem for the smoothed estimators, since these are linear transforms of the residuals lying in the same dimensions as are used to estimate \(\sigma^2\). Thus, the required independence between errors in the estimated parameters and in \(\sigma^2\) is violated. Recomputation of \(\sigma^2\) for each smoothed estimator, based upon only the 597 dimensions orthogonal to it, would be extremely laborious. With a residual subspace of dimension 598, an overlap of one or a few dimensions is probably negligible. Hence, \(\sigma^2\) is assumed to obey a \(\chi^2\) distribution with 598 degrees of freedom, which is independent of the estimated stochastic parameters.
the time history of systematic risk, we conjectured that parameter shifts might have occurred between four dates, chosen to be regularly spaced and to coincide with episodes in economic history, as follows: December 1933, at the heart of the depression; December 1943, midway through World War II; December 1958, near the end of a series of classical business cycles; and December 1973, near the height of uncertainty concerning inflation. Using the procedure in Rosenberg ([1977], eq. (12)), the estimated parameter shifts between successive dates and their standard errors can be computed. The parameter shifts, with t-statistics in parentheses, are as follows: from 1933 to 1943, .047 (.41); from 1943 to 1958, -.335 (-2.96); from 1958 to 1973, .235 (1.88).

The two later intervals roughly coincide with estimated trends in beta, and the cumulative parameter shifts over the intervals are substantial and fairly significant. The first interval includes two apparent uptrends and one downtrend, with little net effect. It misses the very large and highly significant increase in beta from 1929 to 1933.

Despite the fact that the intervals do not overlap, the estimation errors for parameter shifts are correlated. The error correlation between intervals 1 and 2 is -.52; between intervals 1 and 3, -.003; and between intervals 2 and 3, -.53. Strongest correlation occurs between successive intervals, because the terminal value of beta for the first interval is the beginning value for the next. An error in this estimate induces errors of opposite sign in the two adjacent parameter shifts and hence negative correlation.
Let \( P \) denote the \( 3 \times 3 \) error variance matrix for the three-element vector \( \hat{d} \) of estimated parameter shifts. Then, under the null hypothesis that no parameter shifts occurred over the designated intervals, the vector is distributed with mean zero and variance matrix \( \hat{P} \). Hence, a test of the null hypothesis is provided by the statistic \( \hat{d}'\hat{P}^{-1}\hat{d} \), which equals 3.24. This is approximately distributed as \( F(3,598) \).\(^1\) The critical value at the 99 percent level of confidence is 2.62, so constancy over the designated intervals may be rejected.

The pattern of variance and covariances of error in the smoothed estimators gives insight into the smoothing process. Consider, first, estimation error variance. The model specification implies a stationary distribution for \( \delta \), with mean zero and variance \( \Omega \). Thus, in the absence of observations on the return process, the MLE would be zero and the mean square error would be \( \Omega \). The smoothed estimator is able to improve on this estimate by inference from proximate observations of index returns. At the beginning and end points of the history, observations exist at only one side, but in the interior of the history, a two-sided sample exists. Hence, one would expect the standard error to decline toward the interior. This does occur. Standard errors for various observations are as follows. For observations at the beginning of the history: 1, .110; 5, .104; 10, .099; 25, .087. For observations at the end of the history: 600, .103; 595, .093; 590, .086; 575, .086.

\(^1\)As explained in the previous footnote, \( \sigma^2 \) is estimated over the residual subspace of dimension 598, including the three dimensions of the alternative hypothesis.
The standard error at any point in time is also influenced by the information content of proximate observations. The greater the nearby variability of the explanatory variable (the market return), the greater is the precision of the smoothed estimator for beta. As can be seen from close inspection of figure 5, the standard error reaches a low of .069 at $t = 76$, and achieves other local minima at $t = 148$, $t = 243$, and $t = 532$.

Estimation error variance ranges from a maximum of .0121 at $t = 1$ to a minimum of .0048 at $t = 76$. These variances are, respectively, 45% and 18% of the stationary variance $\Omega$, confirming that the smoothed estimators identify most of the variation in the sequential stochastic process.

The smoothed estimate at each point in time can be thought of as the outcome of a two-sided window that places the greatest weight on the nearest observation, with declining weight as temporal distance increases. Two nearby estimates will inevitably exhibit high positive error correlation, because the windows overlap. Figure 6 plots estimation error covariance functions for two cases. First, covariance between $\hat{\delta}_{576|600}$ and $\hat{\delta}_{s|600}$, for $s = 551, \ldots, 500$. Second, error covariance between the "filtered" estimator $\hat{\delta}_{600|600}$ and $\hat{\delta}_{s|600}$, for $s = 550, \ldots, 600$. Covariance declines approximately exponentially with distance in time. Error correlation falls slowly and does not decline to $.5$ until twenty-five to forty months distance.
III.4. Prediction of Systematic Risk

The optimal forecast for beta, defined as the minimum mean square error unbiased prediction, is the sum of \( \hat{\beta} \), \( \delta \), and \( \hat{\epsilon} \). Since the latter component is serially independent with zero mean, the best prediction of that variable is zero. Hence, the prediction problem reduces to forecasting the sequentially varying component. Let \( \tau \) be the forecast horizon, which takes values +1, +2, ... in sequence. Then the optimal predictions, under the assumption that the estimated specification is correct, are given by the recursive formula:

\[
\hat{\beta}_{T+\tau|T} = \hat{\beta} + \hat{\delta}_{T+\tau|T}, \quad \tau = 1, 2, ...
\]

where \( \hat{\delta} \) is given by the recursive formula:

\[
\hat{\delta}_{T+\tau|T} = \phi \hat{\delta}_{T+\tau-1|T}.
\]

This formula is initialized by the filtered estimate, \( \hat{\delta}_{T|T} \).

Let \( \beta^e = \hat{\beta} - \bar{\beta} \) and \( \delta^e_{T+\tau} = \hat{\delta}_{T+\tau} - \delta_{T+\tau} \) denote prediction errors. The variance of forecast error, \( F_{T+\tau|T} \), is given by:

\[
F_{T+\tau|T} = \text{VAR}(\beta^e_{T+\tau|T}) = \text{VAR}(\beta^e) + 2 \text{COV}(\beta^e, \delta^e_{T+\tau}) + \text{VAR}(\delta^e_{T+\tau}).
\]

The necessary variances and covariances are given by the recursive formulas:

\[
\text{COV}(\beta^e, \delta^e_{T+\tau}) = \phi \text{COV}(\beta^e, \delta^e_{T+\tau-1}), \quad \tau = 1, 2, ...
\]

\[
\text{VAR}(\delta^e_{T+\tau}) = \phi^2 \text{VAR}(\delta^e_{T+\tau-1}) + Q.
\]
These formulas are initialized with the properties of the filtered estimate: \( \text{COV}(\tilde{\beta}_T^e, \delta_T^e) \) and \( \text{VAR}(\delta_T^e) \).

As the forecast horizon increases, the beta prediction begins at the filtered estimate for the latest period and approaches the estimated mean value exponentially. The filtered estimate is \( \hat{\beta}_T^e = 1.175 \), and the mean is estimated as \( \hat{\beta} = 1.187 \). Since these two estimates are quite close, the increase in the forecast is not dramatic: after 55 months (the half-life of the sequential process), the forecast will increase by one-half of the difference, or .006.

The increase in error variance is more substantial: from the estimation error variance of .0105 for the filtered estimate, the variance asymptotically approaches a limiting value equal to the sum of the error variance for \( \bar{\beta} \) (.0064) and the stationary variance of the sequential process (.0267). This limit is .0331, corresponding to a standard error of .192, in comparison to the standard error of .103 for the filtered estimate and .105 for the one-period-ahead forecast. The standard error approaches its limit monotonically along a smooth curve—-not exactly exponential. The approach to the limit is more rapid for the standard error than for the estimate itself. For example, after twelve months, the estimate has risen 14 percent of the way to the limit, but the standard error has increased by more than 25 percent.
REFERENCES


