James Ohlson and Barr Rosenberg*
University of California, Berkeley

Systematic Risk of the CRSP Equal-weighted Common Stock Index: A History Estimated by Stochastic-Parameter Regression

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 50-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. Introduction and Summary

Empirical research in finance has long recognized that systematic risks ("betas") are not particularly likely to be stationary. The more sophisticated early work testing the capital asset pricing model viewed the nonstationarity as a practical obstacle to a test of the theory. Much effort was, therefore, expended on attempts to circumvent the problem through the construction of portfolios (see, e.g., Black, Jensen, and Scholes 1972). This approach appears to have some justification. Blume (1971, 1975), for example, concludes that portfolio betas are stable over relatively long periods, but that betas of individual stocks are definitely unstable. However, all of the approaches that have been put forward have an important limitation. The focus has to be almost exclusively on whether stationarity is an adequate hypothesis; to the extent the hypothesis is rejected, the econometric procedures provide little (if any) guidance in terms of what to

*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. Without implicating him, the authors would also like to thank Terry Marsh of the Massachusetts Institute of Technology for an extensive set of very useful comments on an earlier version of the paper. Professor Marsh also directed our attention to a number of recent papers which bear on the topic at hand. These have been included in our references.

(*Journal of Business,* 1982, vol. 55, no. 1)

0021-9398/82/5501-0002$01.50

do. This merely amounts to saying that the aforementioned work does not reflect any real attempt to (parametrically) model and estimate the beta process itself. It is the purpose of the present paper to fill this void in the literature. The basic approach is one of *stochastic-parameter regression,* which permits the statistical identification and estimation of a beta process. Further, it is possible to estimate the beta at any given point in time. (Stochastic-parameter regression methods have long been familiar in statistical theory; see, e.g., Rosenberg [1973].)

The current paper studies the riskiness of a portfolio of equally weighted stocks, and the market portfolio is given by a capitalization-weighted index. As will be evident from the paper, the statistical and computational methodology we develop and implement is completely general; it can be used for any security or group of securities. However, we found it appropriate to use the return on the equal-weighted market portfolio as the dependent variable for two reasons. First, it provides, in a sense, an extreme test of the stationarity hypothesis. If one can reject that betas are stationary in this case, then it is most difficult to believe that individual assets have stationary betas. Second, to minimize estimation error for beta, it is useful to have a dependent variable that is highly correlated with the independent variable. In other words, a low residual variance is equivalent to keeping the "background noise" at a low level.

Besides issues of methodology, there are other reasons why one might want to study the stochastic behavior of the beta of an equal-weighted market index. Of particular interest are the implications suggested by an errors-in-variable perspective on beta estimation, with error arising from misspecification of the market portfolio. At least in a heuristic fashion, the estimation results bear on the consequences of substituting an equal-weighted index for a capitalization-weighted index in beta regressions for other securities. Finally, the question of whether an equal-weighted investment strategy has a positive "alpha" has attracted much attention over the years; in this regard the results are particularly interesting, since the sign of estimated alpha depends on the stochastic model specified.

The results can be summarized very briefly as follows. First, stochastic-parameter regression is both feasible and practical. All parameters can be estimated using maximum-likelihood methods, and confidence regions for the parameters can be developed using a mixture of classical and Bayesian perspectives. Second, the data suggest, rather unambiguously, that the behavior of betas can be attributed to two distinct stochastic factors. There is a tendency for betas to converge rather slowly toward a norm (the stationary mean), and in that sense it is clear that the model of betas requires a "memory"; this is modeled by a stationary first-order, autoregressive process. At the same time, in each period there is a purely random (serially independent) perturbation in beta. Specialized versions of the indicated class of models, such as pure martingale, constant beta, no serial correlation in betas, no purely random perturbation, and so on are all rejected at statistically significant levels.

II. Systematic Risk of the CRSP Equal-weighted Index

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:

$$1 + i_{nt} = (p_{nt} + \text{div}_{nt})/p_{n,t-1}, \tag{1}$$

where $1 + i$ is total return, i is the "rate of return," p_{ns} is stock price at the end of month s, and div_{nt} is dividends paid in month t. (Prices and dividends must be computed on the basis of comparable shares 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 stock returns can be defined as $\Sigma_n w_{nt}(1 + i_{nt})$. 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_{Mnt} is the market weight of each common stock at time t, defined as w_{Mnt} = (market value of outstanding common stock n)/(aggregated market value of all stocks), then the market rate of return is defined by:

$$1 + i_{Mt} = \sum_{n=1}^{N} w_{Mnt}(1 + i_{nt}). \tag{2}$$

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) at the University of Chicago; this capitalization-weighted index of all common stocks listed on the New York Stock Exchange is 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's 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:

$$1 + i_{Et} = \sum_{n=1}^{N} \frac{1}{N} (1 + i_{nt}). \tag{3}$$

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.[1] 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, and many researchers continued to use EW even after MW became available.

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:

$$r_{nt} = (1 + i_{nt}) - (1 + i_{Ft}) = i_{nt} - i_{Ft}, \tag{4}$$

where i_{Ft} is the rate of return on a riskless asset (taken as the annualized return on 90-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:

$$r_{nt} = \alpha_{nt} + \beta_{nt} r_{Mt} + u_{nt}, \tag{5}$$

where β_{nt} is the "systematic risk coefficient," commonly known as beta; α_{nt} is the "abnormal return," which is zero under the simplest version of the capital asset pricing model (CAPM); 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:

$$r_{nt} = \ln(1 + i_{nt}) - \ln(1 + i_{Ft});\ r_{Mt} = \ln(1 + i_{Mt}) - \ln(1 + i_{Ft}). \tag{4'}$$

1. E.g., 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.

This logarithmic form of equation (4), to be used in the present study, has advantages and disadvantages; these are discussed in Rosenberg and Marathe (1978, sec. II.A.3).

The equal-weighted logarithmic excess return is defined analogously:

$$r_{Et} = \ln(1 + i_{Et}) - \ln(1 + i_{Ft}). \tag{6}$$

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:

$$r_{Et} = \alpha(t) + \beta(t)r_{Mt} + v_t, \tag{7}$$

where α and β are written as functions of time, since they may not be constant. If $\alpha(t)$ is identically zero, $\beta(t)$ is constant, and v_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 v_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 α. 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, $\mathrm{var}(v_t)$, of an equal-weighted strategy. Hence, they bear directly on the desirability of such a strategy.

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, one would then have to ask, and answer, what "type" of stochastic evolution the parameters are "most likely" to follow. Existing theory, unfortunately, provides only a very modest amount of information that bears on the question. In addition, since the market weights are stochastic, it is readily seen that it is practically impossible to derive the stochastic behavior of α, β in (7) as a function of the behavior of individual securities.

However, whatever the economic determinants of the stochastic behavior of the parameters are, one might hypothesize that the effects on the parameters can be captured (or "summarized") by a few stochastic variables. In other words, one simply tries to develop a model that is computationally feasible and parsimonious and, yet, at the same time, can capture whatever permanent and/or transitory effects operate on the parameters. This leads naturally to the following approach.

First, the model should allow for a memory that admits regression toward the norm as a special case. This is captured by a second-order autoregressive process, with parameters ϕ and ϕ':

$$\beta_t = b_0 + \phi\beta_{t-1} + \phi'\beta_{t-2} + d_t, \tag{8}$$

where d_t is a random shift which is serially independent, with mean zero in all periods and variance Q. The model may be rewritten, provided that $\phi + \phi' < 1$, $\phi' - \phi < 1$, and $|\phi'| < 1$, in the form:

$$\beta_t = \bar{\beta} + \phi(\beta_{t-1} - \bar{\beta}) + \phi'(\beta_{t-2} - \bar{\beta}) + d_t, \tag{9}$$

where $\bar{\beta} = b_0/(1 - \phi - \phi')$. In this form, the process is convergent with stationary mean value $\bar{\beta}$, and with deviations $\delta_t \equiv (\beta_t - \bar{\beta})$ obeying a second-order autoregressive process:

$$\delta_t = \phi\delta_{t-1} + \phi'\delta_{t-2} + d_t. \tag{9'}$$

As another special case, when $b_0 = \phi' = 0$ and $\phi = 1$, the model produces a random walk with zero drift:

$$\beta_t = \beta_{t-1} + d_t. \tag{10}$$

Second, the tendency for stochastic drift in beta does not exhaust the possible forms of stochastic variation. One might hypothesize that not all factors which impart a change on beta must also have a "carry-forward" effect: Equation (9) implies that, in general, any realization d_t affects $\beta_{t+\tau}$, where τ is any nonnegative integer. The process is, thus, highly restrictive in that it precludes the existence of events which have a purely temporary, nonpermanent, impact on beta. Such possibilities are not difficult to imagine, since betas might be determined partially by unexpected changes in whatever factors happen to be relevant. For a more complete discussion of these points, see Rosenberg (1974) and Rosenberg and Guy (1976).[2] Hence, beta in each period is allowed to exhibit a purely random increment, ϵ, with mean zero and stationary variance Λ. The model including this phenomenon may be written:

$$r_{Et} = \alpha + (\beta_t + \epsilon_t)r_{Mt} + v_t, \tag{11}$$

where β_t is determined by (8). Alternatively, (11) can be specified as:

$$r_{Et} = \alpha + \beta_t r_{Mt} + u_t,\ u_t = \epsilon_t r_{Mt} + v_t, \tag{12}$$

where $E(u_t) = 0$; $\text{var}(u_t \mid r_{Mt}) = \sigma^2 + r_{Mt}^2\Lambda$; $\text{cov}[u_t,u_s] = 0$ for $t \neq s$.

The above formulation, (12) combined with (8), thus implies that the value of beta at time t can be viewed as a function of two different sources of perturbations: There is a sequence of factors d_t, d_{t-1}, d_{t-2}, . . . having a persistent impact, and there is a purely temporary factor, ϵ_t. Assuming that these factors have positive variances, ϵ_t, d_t combine into one source of uncertainty, $\epsilon_t + d_t$, if and *only* if, $\phi = \phi' = 0$.

It should be noted, however, that there are other reasons why (12) with $\Lambda > 0$ may be an appropriate specification. Specifically, if the variance of u_t is not constant and positively correlated with nonstationary $\text{var}(r_{Mt})$, then the disturbance term is heteroscedastic and the specification $\text{var}(u_t \mid r_{Mt}) = \sigma^2 + r_{Mt}^2\Lambda$ is one way of modeling the phenomenon. The data are necessarily noninformative concerning the distinction between heteroscedasticity related to $\text{var}(r_{Mt})$ and temporary shocks to beta. Even so, it is useful to think of beta as being a function of two sources of perturbation, where the permanent component of beta (β_t), and its process, is considered with particular interest.

It should further be mentioned that the omission of variables, whose squared residual, after regression on r_{Mt}, is correlated with r_{Mt}^2 can give rise to an estimated Λ which is positive. The omitted effects become a component of the disturbance which appears to be heteroscedastic in relation to r_{Mt}^2. The literature does not seem to suggest important problems of this kind,[3] and we hope that the beta process modeled captures the significant features of the economic reality.

The stochastic behavior of alpha can be considered from a similar point of view. In addition, capital market theory does, of course,

2. The theory of beta determination in dynamic environments has been considered by Garman and Ohlson (1980) and Ohlson and Garman (1980).

3. Of course, to the extent that r_{Mt} is regarded as an incorrect measure of the "true" market factor(s) (or factor portfolio[s]), then "all bets are off." Unless the relevant measures are available, the only way one can deal with this issue is to model the error process and its correlations to other stochastic components. These tasks are much beyond the scope of the paper, and here it suffices to mention that implications of such misspecifications appear to be relatively difficult to identify. The point is applicable not only with respect to the beta process but also the alpha process, considered below.

suggest that α_t should take on certain values, and this might be of some aid in the analysis.

However, the issues involved with alpha are of limited interest in the estimation problem at hand. As discussed below, on prior grounds it appears unlikely that one can successfully identify nontrivial nonconstancy in alpha through stochastic-parameter models that do not rely on concomitant variables which explain the change in alpha.

Preliminary Simplification

Before analyzing the equal-weighted index, we applied the methodology to the history of returns on an individual common stock. The use of a dependent variable other than r_E as a basis for this analysis provides an extraneous basis for model selection, so that test statistics in our application can be taken at face value. The complete model includes seven "specification parameters": ϕ, ϕ', and Q for the beta process; three analogous parameters for the alpha process; and the variance, Λ, of the random increment to beta. We found that the data were uninformative concerning the second-order coefficient ϕ' 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, var(u_t), is very large: For a typical asset, the standard deviation of u_t is roughly .08 per month. A change in alpha as substantial as .005 per month would, therefore, contribute less than 1/200 of unexplained squared return in a month. Hence, a change in alpha disappears in the background noise unless it persists for hundreds of months. Consequently, a stochastic process for alpha is completely obscured. For comparison, a moderate 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, which is roughly 10 times as great as the hypothesized effect due to change in α, is larger in comparison to residual return, but still cannot emerge distinctly until several years have elapsed. 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 affect the probability of the fitted beta history. This suggests that a first-order process should suffice.

We anticipated that data for the equal-weighted index would also be essentially uninformative concerning these parameters. Index residual variance is roughly one-third as great as the typical stock, but is still sufficient to obscure substantial changes in alpha and all rapid transients in the beta process. We therefore chose to fit a simplified model, omitting the poorly distinguished parameters, in which α is constant and β follows a first-order process:

TABLE 1 Various Transforms of the Specification Parameters

Natural	Relative Variances	Unrestricted	Transformed
ϕ	$\phi = \theta_1$	$\phi = z_1^2/(1 + z_1^2)$	ϕ
Q	$Q = \sigma^2\theta_2$	$Q = \sigma^2 z_2^2$	$\rho = \phi\Omega/(\Omega + \Lambda)$
Λ	$\Lambda = \sigma^2\theta_3$	$\Lambda = \sigma^2 z_3^2$	$\Lambda = \sigma^2\Lambda^*$

$$
\begin{aligned}
r_{Et} &= \alpha + \bar{\beta} r_{Mt} + \delta_t r_{Mt} + u_t \qquad && t = 1, \ldots, 600 \\
\delta_t &= \phi\delta_{t-1} + d_t && t = 2, \ldots, 600 \qquad (13)
\end{aligned}
$$
$$E(u_t) = 0;\ \mathrm{var}(u_t) = \sigma^2 + r_{Mt}^2\Lambda;\ E(d_t) = 0;\ \mathrm{var}(d_t) = Q.$$

The model has three specification parameters: ϕ, Q, and Λ. When $|\phi| < 1$, the beta process is stationary; $\bar{\beta}$ is the expected value for β_t; and the stationary variance of δ_t is:[4]

$$\mathrm{var}(\delta_t) \equiv \Omega = Q/(1 - \phi^2). \qquad (14)$$

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 δ is a stationary process, initialization consists in applying the stationary distribution as the prior distribution for δ_1:

$$E(\delta_1) = 0;\ \mathrm{var}(\delta_1) = \Omega. \qquad (15)$$

If, on the other hand, $\phi = 1$, then the process is a driftless random walk, and the distribution for β 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.

III. Estimating the Stochastic Specification

Methodology

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

In the transformation to $\underset{\sim}{\theta}$, the variances Q and Λ are expressed in relative terms, as multiples of the disturbance variance σ^2. This causes

4. The stationary variance for δ is found by observing that $\mathrm{var}(\delta_t) = \phi^2\mathrm{var}(\delta_{t-1}) + \mathrm{var}(d_t)$ and solving for $\Omega = \mathrm{var}(\delta_t) = \mathrm{var}(\delta_{t-1})$.

σ^2 to become the scale parameter for all second moments in the model, so that it enters linearly in the variance matrix of the data vector r_E. The matrix (of dimension 600) may then be written as $\sigma^2\underset{\sim}{V}(\underset{\sim}{\theta})$. 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 $\underset{\sim}{\theta}$ only, $l(\underset{\sim}{\theta})$, such that l is easily computed for given $\underset{\sim}{\theta}$.[5]

The parameter vector, $z = (z_1:z_2:z_3)'$, is used to define the specification during maximum-likelihood estimation. As the vector $\underset{\sim}{z}$ ranges over three-dimensional Euclidean space, the specification varies over the admissible region: $\underset{\sim}{0} \leqq \phi < 1$; $\underset{\sim}{0} \leqq \underset{\sim}{Q}$; $\underset{\sim}{0} \leqq \Lambda$. Hence, unconstrained maximization over $R(\underset{\sim}{z})$ corresponds to a constrained maximization over $R(\underset{\sim}{\theta})$. Thus, by use of $\underset{\sim}{z}$, optimization programs for unconstrained problems may be used.[6]

Finally, in the "transformed" parameter vector, Q is replaced by the first-order autocorrelation coefficient, ρ, of total systematic risk. The coefficient is defined in terms of Ω (the stationary variance of the sequential beta process), Λ (the variance of the random process) and ϕ (the convergence rate parameter); specifically, one gets:[7]

$$\rho = \mathrm{corr}(\bar{\beta} + \delta_t + \epsilon_t, \bar{\beta} + \delta_{t-1} + \epsilon_{t-1}) = \frac{\phi\Omega}{\Lambda - \Omega}. \tag{16}$$

Also, the variance of the random beta increment, Λ, is expressed relative to disturbance variance by Λ^*. 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 the subsection on the shape of the likelihood function below).

The estimation procedure is based on the decomposition of l. For given $\underset{\sim}{\theta}$, the stochastic-parameter regression problem may be shown to

5. In other words, the maximum of the log-likelihood function is obtained by a two-step procedure:

$$\max_{\theta,\sigma^2,\alpha,\bar{\beta}} l = \max_{\underset{\sim}{\theta}} \{ \max_{\sigma^2,\alpha,\bar{\beta}} l\,(./\underset{\sim}{\theta})\} \equiv \max_{\underset{\sim}{\theta}} l(\theta),$$

where the first (conditional) maximization is solved by a GLS procedure (see below), and the second maximization is solved by a nonlinear optimization program.

6. The transformation from $\underset{\sim}{z}$ to θ 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.

7. The derivation of ρ is, thus, as follows:

$$\begin{aligned}\rho &\equiv \mathrm{cov}(\epsilon_t + \beta_t, \epsilon_{t-1} + \beta_{t-1})/\mathrm{var}(\epsilon_t + \beta_t)\\ &= \mathrm{cov}(\beta_t,\beta_{t-1})/[\mathrm{var}(\epsilon_t) + \mathrm{var}(\beta_t)] = \phi\ \mathrm{var}(\beta_t)/[\mathrm{var}(\epsilon_t) + \mathrm{var}(\beta_t)]\\ &\equiv \phi\Omega/(\Lambda + \Omega).\end{aligned}$$

be equivalent to a generalized least-squares (GLS) regression, with the variance matrix for the data, $\underset{\sim}{V}(\underset{\sim}{\theta})$, dependent on $\underset{\sim}{\theta}$. The problem, then, is to evaluate the familiar likelihood function for a GLS regression, which involves computing the inverse of V (a 600×600 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 l, which diagonalizes the variance matrix. Thus, l (eq. [29] in Rosenberg [1973]) is evaluated by a set of recursive formulas (simplifications of eqq. (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 $\underset{\sim}{V}$ would require approximately 36,000 times as many computations.

The estimation procedure of $\underset{\sim}{\theta}$ 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, chap. 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.

Estimates of the Specification for Beta

Table 2 reports maximum-likelihood estimates (MLEs) for six possible model specifications. Model 1 is the complete model (eq. [13]), with an unrestricted first-order process, plus an additive random increment. The MLE for the convergence factor, ϕ, 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 55 months. The variance Q of the shifts is .000655, while the variance Λ of the random increment ϵ_t is quite large: .0839.

Model 2 is restricted to $\phi = 1.0$ and therefore implies a driftless random walk for β_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 χ^2 with one degree of freedom. The critical value at the 99.5% level is 7.88. $2\,(l_{\max} - l) = 11.38$ exceeds the critical value, so the driftless random-walk hypothesis may be rejected at the 99.5% level of confidence. Model 2 is, thus, significantly different from model 1, in spite of the fact $\phi = .9875$ would seem to be "almost equivalent" to $\phi = 1$. The intuition behind the observation is that it is misleading to view the two values of ϕ as almost equivalent; a half-life of 55 months, although quite long, is quite distinct from a half-life of infinite duration. (Alternatively, convergent processes are quite distinct from nonconvergent processes even when the convergence rate is relatively slow.)

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, ϵ_t, also present 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 Q to zero also in model 5.[8] Thus, model 5 admits only random variation in beta. As ϕ declines toward zero, l 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, $\underset{\sim}{\theta} = \underset{\sim}{0}$, and the model specifies a simple linear regression. There are three restrictions on $\underset{\sim}{\theta}$, so twice the deficiency in log likelihood is asymptotically distributed as χ^2 with three degrees of freedom. The critical value at the 99.5% level of confidence is 12.84; twice the log-likelihood deficiency is 205.74, so beta constancy is overwhelmingly rejected.[9]

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 Ω given in (14). The random-walk model is nonstationary, and no variance is defined. Estimated Λ increases as Ω 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 S.D. of .33). This accounts for most of the large increase in l 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.

The latter process accounts for only a fraction of total beta variance (.0267/.1106 = .241), but it is nevertheless of great interest for two reasons: First, it is forecastable; second, it is persistent over time. 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% of beta variance in

8. Alternatively, one can put $\Lambda = 0$ and estimate Q.

9. A model with sequential variation in beta only ($\Lambda = 0$) performs worse than model 5, but significantly better than model 6.

TABLE 2 **Maximum Likelihood Estimates of Specification Parameters**

	Convergence Rate	Variance of Shift	Random Variance	Deficiency in Log Likelihood
Model for Beta	ϕ	Q	Λ	$l_{\max} - l$
Sequential + random variation	.9875	.000665	.0839	0
Driftless (ϕ = 1.0) + random	set at 1.0	.000420	.0847	5.67
Sequential (ϕ = .8) + random	set at .8	.0149	.0668	6.21
Sequential (ϕ = .5) + random	set at .5	.0520	.0364	8.44
Random variation only	set at 0	set at 0	.1030	13.59
Constant	set at 0	set at 0	set at 0	102.87

TABLE 3 **Components of Beta Variance and Variance of r_E**

	Variance of Beta			Variance of r_E			
Model	Ω	Λ	Total	$\bar{\beta}^2\text{var}(r_M)$	$\Omega E(r_M^2)$	$\Lambda E(r_M^2)$	σ^2
1	.0267	.0839	.1106	.004838	.000093	.000294	.000311
2	nonstationary	.0847	n.a.	n.a.	n.a.	.000297	.000311
3	.0415	.0668	.1083	.005036	.000145	.000234	.000312
4	.0694	.0364	.1058	.005089	.000243	.000128	.000316
5	0	.1030	.1030	.005143	0	.000361	.000322
6	0	0	0	.005542	0	0	.000689

a single month, but accounts for more than 75% of interyear variance in average annual beta.

It is also interesting that the first-order autocorrelation coefficient, ρ, 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 ϕ is nearly offset by the increase in Q.

The difference between the equal-weighted index return and its mean may be written as follows: $r_{Et} - E(r_{Et}) = \alpha + (\bar{\beta} + \delta_t + \epsilon_t)r_{Mt} + u_t - E[\alpha + (\bar{\beta} + \delta_t + \epsilon_t)r_{Mt} + u_t]$. Since α and $\bar{\beta}$ are constants, and $E(\delta_t) = E(\epsilon_t) = E(u_t) = 0$, and the stochastic beta elements δ_t and ϵ_t are uncorrelated with the market return, this becomes:

$$= \bar{\beta}(r_{Mt} - E[r_{Mt}]) + (\delta_t + \epsilon_t)r_{Mt} + u_t. \tag{17}$$

Hence, the variance of r_E is:

$$\begin{aligned} \text{var}(r_E) &= \bar{\beta}^2\, \text{var}(r_M) + \text{var}[\delta_t] + \text{var}[\epsilon_t])E(r_M^2) + \text{var}(u_t) \\ &= \bar{\beta}^2\, \text{var}(r_M) + (\Omega + \Lambda)E(r_M^2) + \sigma^2. \end{aligned} \tag{18}$$

When $\hat{\bar{\beta}}$ (table 4, below) is substituted into (18), the decomposition of variance of r_E in the second panel of table 3 results. Total sample variance of r_E is .00623. The interpretation of the results for model 1 is that 77.7% of the variance of r_E is attributable to the normal value of systematic risk, 1.49% to stationary sequential variation in beta, 4.71% to random fluctuation in beta, and 4.99% to residual variance.

The decomposition of the variance of r_E is not comprehensive by construction, and the sum of the percentages is not 100. Lengthy but straightforward derivations would demonstrate that the missing term reflects the sample correlation between β_t and realized r_{Mt}^2. In the model specified, the *expectation* of this interaction is zero, but, in the sample history described below, beta is seen to be much higher during the depression years when the market variance is also larger. Thus, the balance of 11% of experienced variance is attributed to this positive correlation in the sample. Of course, it could also be interpreted as an indication of model misspecification. In the absence of a sampling theory of the decomposition, it is not possible, however, to confirm that the 11% deviation is statistically significant; because of the long half-life of the estimated process and the fact that almost all extreme values of r_{Mt} cluster in an 8-year interval during the depression, the interaction could occur due to a chance increase in beta during that period. On the other hand, the greater uncertainty of economic activity during the depression could have induced greater relative exposure in smaller companies and hence *caused* the increase in beta during the period.

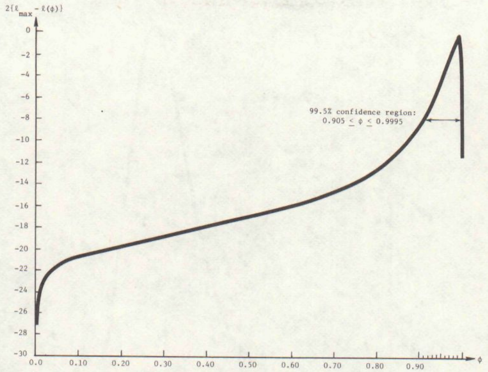

FIG. 1.

The Shape of the Likelihood Function or Bayesian Posterior

The convergence rate ϕ is of special interest, since the model ranges from purely random beta through a stationary autoregressive process to a driftless random walk as ϕ 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 ϕ is defined by the range within which l falls short of l_{max} by less than the critical value. The 99.5% confidence region is shown on the figure: It is the interval within which $2(l_{max} - l(\phi)) \leqq \chi^2_{.005} = 7.87$. 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]). In this problem, the correction factor is relatively insignificant.[10] The conditional posterior density for ϕ, given the MLE for Q and Λ, is shown in figure 2. The mode of the Bayesian posterior effectively coincides with the MLE.

From table 2 it is apparent that, as ϕ decreases, the constrained

10. This was verified computationally for some specific cases.

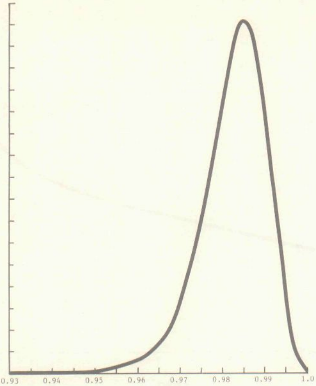

FIG. 2.—A conditional posterior distribution for ϕ: $P(\phi \mid Q = .000665, \Lambda = .0839)$.

maximum is obtained at larger values of Q and smaller values of Λ. To underscore this point, a pair of conditional densities for Q and another pair for Ω are shown in figures 3 and 4. For Q, the posterior densities are plotted for two values of ϕ that are relatively close to one another: The MLE $\hat{\phi} = .9875$, $\phi = .999$. (In each case, the chosen value of Λ is the MLE, given ϕ.) Although ϕ 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 Λ, posterior densities are shown for two extremely different models: first, model 1 with MLE for ϕ and Q; second, model 5 with $\phi = Q = 0$. The densities are slightly skewed but more nearly symmetric than the conditional densities for ϕ 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 ϕ to Q and Λ would render the quadratic approximation invalid. To check this, we fitted a quadratic form in ϕ, Q, and Λ, with extremum 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

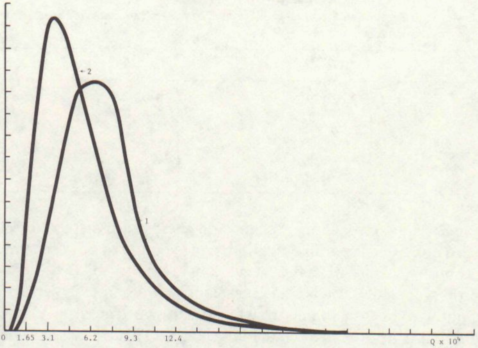

FIG. 3.—Two conditional posterior distributions for Q: (1) $P(Q \mid \phi = .9875, \Lambda = .0839)$; (2) $P(Q \mid \phi = .999, \Lambda = .0846)$.

the 90% asymptotic confidence region for the parameters. Two hundred and twenty-five such evaluations of θ, $l(\theta)$ were obtained. The least-squares fit of the quadratic model achieved an R^2 of only .760, confirming that the quadratic approximation of the log-likelihood function was poor indeed. The result thus confirms what one possibly might have conjectured on the basis of the shapes of the Bayesian posterior distributions in figures 2–4; the log-likelihood function is not nearly quad-

FIG. 4.—Two conditional posterior distributions for Λ: (1) $P(\Lambda \mid \phi = .9875, Q = .000665)$; (2) $P(\Lambda \mid \phi = 0, Q = 0)$.

TABLE 4 Regression Parameter Estimates and SEs Implied by Three Alternative Models

Model	$\hat{\alpha}$	$\hat{\bar{\beta}}$	$\hat{\beta}_{601/600}$
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)

ratic, and, as a consequence, reliance on asymptotic theory to estimate standard error of the parameter estimates is bound to be somewhat less than robust (in spite of 600 observations).

We then considered whether a transformed parameter vector might allow improved fit. Only one transformation was tried: replacement of Q by ρ, and Λ by Λ^*, 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, $(\Omega + \Lambda)$ might replace Λ^*.

IV. Regression Parameter Estimates

Estimates of the Unknown Parameters

Each model specification $\underset{\sim}{\theta}$ implies a variance matrix $\underset{\sim}{V}(\underset{\sim}{\theta})$ for the observations, and a generalized least-squares regression, in which the regression coefficients are α and $\bar{\beta}$, and the "scale parameter" for the variances and covariances of disturbances is σ^2. Using $\hat{\sigma}^2$ and $\underset{\sim}{V}(\hat{\underset{\sim}{\theta}})$, the standard errors of the coefficients may be computed.[11] 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.

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 variable with coefficient α is nearly orthogonal

11. From a Bayesian standpoint, it would be preferable to compute the conditional distribution of these parameters for each $\underset{\sim}{\theta}$ and then integrate with respect to the posterior density of $\underset{\sim}{\theta}$ 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 $\hat{\alpha}$, $\hat{\bar{\beta}}$, and $\hat{\beta}$, conditioned on the presumption that the MLE specification of θ is the correct one.

to the market-return variable. When two variables are orthogonal, change of the variance matrix in the dimension of one variable affects the computed standard error of the other little, if at all). 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% in the annual rate of return (40 basis points). Compounded over the 50-year data history, this becomes a difference of more than 20% 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 (Fisher 1966, pp. 197–99). 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. Of course, nobody would seriously argue that a beta for a single security (or portfolio) could be (approximately) constant over a 50-year time span. Nevertheless, the common procedure of estimating a constant beta model using data over successive intervals is not only somewhat arbitrary, but also statistically inefficient. These kinds of procedures can never be quite statisfactory, since they deal with the nonstationary problem through data elimination.

The last column in table 4 compares predictions for beta in period 601. The subscript $T + 1 \mid T$ denotes a prediction based on information through period T, with $T = 600$ in this case. These predictions are based on 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)–(22) below. Notice the substantial differences in forecasts. The increased standard error in model 1 stems from imperfect prediction of δ_{T+1}.[12]

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, $\hat{\beta}_{s|T} = \hat{\bar{\beta}} + \hat{\delta}_{s|T}$, subscripted $s \mid T$ 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% confidence region.[13]

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 the time history of systematic risk, we conjecture 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

12. It should be recalled that the estimated SEs presume each of the three specifications to be, in turn, correct. Thus, the increases in the forecasting errors must not be interpreted as an ordering of the models' forecasting properties.

13. 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 σ^2 of second moments of stochastic terms. This raises an additional problem for the smoother estimators, since these are linear transforms of the residuals lying in the same dimensions as are used to estimate σ^2. Thus, the required independence between errors in the estimated parameters and σ^2 is violated. Recomputation of σ^2 for each smoothed estimator, based on 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, $\hat{\sigma}^2$ is assumed to obey a χ^2 distribution with 598 degrees of freedom, which is independent of the estimated stochastic parameters.

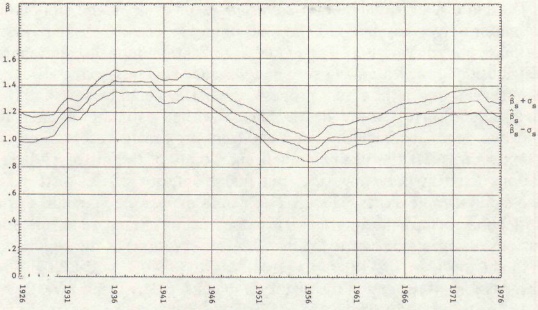

FIG. 5.—Estimated history ± SE for systematic risk.

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, $-.53$. The 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 $\underset{\sim}{P}$ denote the 3×3 error-variance matrix for the three-element vector $\underset{\sim}{d}$ 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 $\underset{\sim}{P}$. Hence, a test of the null hypothesis is provided by the statistic $\underset{\sim}{d}'\underset{\sim}{P}^{-1}\underset{\sim}{d}$, which equals 3.24. This is approximately distributed as $F(3,598)$.[14] The critical value at the 99% level of confidence is 2.62, so constancy over the designated intervals may be rejected.

14. As explained in the previous footnote, σ^2 is estimated over the residual subspace of dimension 598, including the three dimensions of the alternative hypothesis.

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 δ, with mean zero and variance Ω. Thus, in the absence of observations on the return process, the MLE would be zero and the mean square error would be Ω. 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: $t = 1$ (.110); $t = 5$ (.104); $t = 10$ (.099); $t = 25$ (.087). For observations at the end of the history: $t = 600$ (.103); $t = 595$ (.093); $t = 590$ (.086); $t = 575$ (.086).

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, Ω, which confirms 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 $s = 551, \ldots, 600$. 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 25–40 months distance.

Prediction of Systematic Risk

The optimal forecast for beta, defined as the minimum mean-square-error unbiased prediction, is the sum of $\hat{\beta}$, $\hat{\delta}$, and $\hat{\epsilon}$. Since the last 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 τ be the forecast

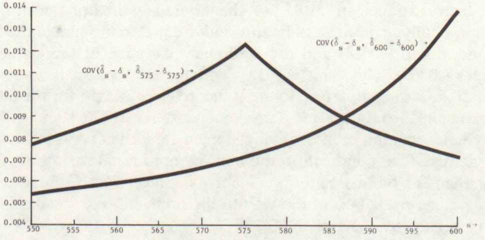

FIG. 6.—Estimation-error covariance

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 formula

$$\hat{\beta}_{T+\tau\mid T} = \hat{\bar{\beta}} + \hat{\delta}_{T+\tau\mid T}, \qquad \tau = 1, 2, \ldots, \tag{19}$$

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

$$\hat{\delta}_{T+\tau\mid T} = \phi\hat{\delta}_{T+\tau-1\mid T}. \tag{20}$$

This formula is initialized by the filtered estimate, $\hat{\delta}_{T\mid T}$.

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

$$F_{t+\tau\mid T} \equiv \mathrm{var}(\hat{\beta}_{T+\tau\mid T} - \beta_{T+\tau}) = \mathrm{var}(\bar{\beta}^e) + 2\,\mathrm{cov}(\bar{\beta}^e,\delta^e_{T+\tau}) + \mathrm{var}(\delta^e_{T+\tau}). \tag{21}$$

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

$$\begin{aligned} \mathrm{cov}(\bar{\beta}^e,\delta^e_{T+\tau}) &= \phi\,\mathrm{cov}(\bar{\beta}^e,\delta^e_{T+\tau-1}) \qquad \tau = 1, 2, \ldots \\ \mathrm{var}(\delta^e_{T+\tau}) &= \phi^2\mathrm{var}(\delta^e_{T+\tau-1}) + Q. \end{aligned} \tag{22}$$

These formulas are initialized with the properties of the filtered estimate: $\mathrm{cov}(\bar{\beta}^e,\delta^e_T)$ and $\mathrm{var}(\delta^e_T)$.

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 $\hat{\beta}_{T\mid T} = 1.175$, and the mean is estimated as $\hat{\bar{\beta}} = 1.187$. 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, which corresponds to a standard error of .182, 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 12 months, the estimate has risen 14% of the way to the limit, but the standard error has increased by more than 25%.

References

Ansley, C. F. 1980*a*. Signal Extraction in Finite Series and the Estimation of Stochastic Regression Coefficients. *Proceedings of the American Statistical Association,* Business and Economic Statistics Section, pp. 251–60.

Ansley, C. F. 1980*b*. A Note on Grenanden Conditions for the GLS Model With Unknown Covariance Matrix. Working Paper, Graduate School of Business, University of Chicago.

Ansley, C. F., and Newbold, P. 1980. Finite Sample Properties of Estimators for Autoregressive Moving Average Models. *Journal of Econometrics* 13, no. 2 (June): 159–84.

Black, F., and Scholes, M. 1974. From Theory to a New Financial Product. *Journal of Finance* 29, no. 2 (May): 399–412.

Black, F.; Jensen, M. C.; and Scholes, M. 1972. The Capital Asset Pricing Model: Some Empirical Tests. In M. C. Jensen (ed.), *Studies in the Theory of Capital Markets*. New York: Praeger.

Blume, M. 1971. On the Assessment of Risk. *Journal of Finance* 26, no. 1 (March): 1–10.

Blume, M. 1975. Betas and Their Regression Tendencies. *Journal of Finance* 30, no. 3 (June): 785–95.

Brent, R. P. 1972. *Algorithms for Minimization without Derivatives*. Englewood Cliffs, N.J.: Prentice-Hall.

Brown, P.; Kleidon, A. W.; and Marsh, T. 1980. Size Related Anomalies in Stock Returns. Working Paper, Graduate School of Business, University of Chicago.

Cooley, T. F.; Rosenberg, B.; and Wall, K. 1977. A Note on Optimal Smoothing for Time Varying Coefficient Problems. *Annals of Economic and Social Measurement* 6, no. 4 (Fall): 453–56.

Fisher, L. 1966. Some New Stock Market Indexes. *Journal of Business* 39, no. 1 (January): 191–225.

Fisher, L., and Lorie, J. H. 1964. Rates of Return on Investments in Common Stocks. *Journal of Business* 37, no. 1 (January): 1–21.

Garman, M., and Ohlson, J. 1980. Information and the Sequential Valuation of Assets in Arbitrage-free Economies. *Journal of Accounting Research* 18, no. 2 (Autumn): 420–40.

Jensen, M. C. 1972. Capital Markets: Theory and Evidence. *Bell Journal of Economics and Management Science* (Autumn), pp. 357–98.

Miller, H., and Scholes, M. 1972. Rates of Return in Relation to Risk: A Re-examination of Some Recent Findings." In M. C. Jensen (ed.), *Studies in the Theory of Capital Markets*. New York: Praeger.

Ohlson, J., and Garman, M. 1980. A Dynamic Equilibrium for the Ross Arbitrage Model. *Journal of Finance* 35, no. 3 (June): 675–84.

Rosenberg, B. 1973. Random Coefficient Models: The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression. *Annals of Economic and Social Measurement* 2, no. 4 (October): 399–428.

Rosenberg, B. 1974. Extra-Market Components of Covariance in Security Returns. *Journal of Financial and Quantitative Analysis* 9, no. 2 (March): 263–74.
Rosenberg, B. 1977. Estimation Error Covariance in Regression with Sequentially Varying Parameters. *Annals of Economic and Social Measurement* 6, no. 4 (September): 457–62.
Rosenberg, B., and Guy, J. 1976. Prediction of Beta from Investment Fundamentals. Part 1. *Financial Analysts Journal* 32, no. 3 (May/June): 60–72.
Rosenberg, B., and Guy, J. 1976. Prediction of Beta from Investment Fundamentals. Part 2. *Financial Analysts Journal* 32, no. 4 (July/August): 62–70.
Rosenberg, B., and Marathe, V. 1978. Tests of Capital Asset Pricing Hypothesis. In H. Levy (ed.), *Research in Finance,* Greenwich, Conn.: JAI.