MEASURING NONSTATIONARITY IN THE PARAMETERS OF THE MARKET MODEL

Melvin J. Hinich and Richard Roll

I. INTRODUCTION

Nonstationarity is an important and pervasive problem in the applications of financial theory. The industrial clients of modern finance theory, such as investment advisory firms, frequently apply statistical techniques to time series that extend over many periods. Virtually all analyses of time series are subject to a nonstationarity problem. There is usually no theoretical reason nor practical guarantee that the data used for estimation were generated by a process whose parameters are stationary.

The purpose of this chapter is to apply a simple technique for measuring nonstationarity to the one-factor market model (MM1) of asset returns,

$$R_{j,t} = \alpha_{j,t} + \beta_{j,t}R_{m,t} + \varepsilon_{j,t} \tag{1}$$

Research in Finance, Volume 3, pages 1–51

ISBN: 0-89232-218-7

in which $R_{j,t}$ is the observed rate of return on asset j during time period t, $R_{m,t}$ is the return on a market index, $\beta_{j,t}$ is a parameter which measures the risk of asset j, $\alpha_{j,t}$ is another parameter which can be given several interpretations, and $\varepsilon_{j,t}$ is a stochastic disturbance term with zero mean.

MM1 is related to the theory of asset price determination, the capital asset pricing model (CAPM) enunciated first by Sharpe (1974) and Lintner (1965) who derived market equilibrium conditions based on the assumption that all investors followed Markowitz's (1952) advice to diversify their investments. In the original CAPM, there were only two periods, so all parameters were stationary (trivially). The parameter α_j was specified as $\alpha_j = (1 - \beta_j)R_F$ where R_F was a riskless rate of interest. In the later model of Black (1973), however, there was no riskless borrowing and α_j became related to the mean return on a portfolio whose return was orthogonal to $R_{m,t}$.

Since excellent expositions of the CAPM and MM1 are available elsewhere,[1] we will not give yet another one. Suffice it note that the original CAPM, or closely related variants such as MM1, are widely applied in the financial industry. In portfolio management, there are probably no more important current paradigms. They aid decision making in risky environments. They have been used for assessing portfolio performance, for measuring the impact of various events on security prices,[2] for measuring rates of discount to use in the valuation of uncertain cash flows, and for other purposes.

MM1 also can be considered a special case of a generating mechanism that leads to the Ross (1976) Arbitrage Pricing Theory (APT). This theory has recently become prominent as a very general model of intertemporal asset price variation and expected return equilibrium.[3]

The APT begins with an *assumption* that returns are generated linearly by a limited set of stochastic factors. MM1 imposes two additional assumptions: (*a*) that the number of factors is exactly one; and (*b*) that this single factor can be measured by some market index. By allowing the parameters of MM1 to change over time as in specification (1), we effectively relax these two requirements. The parameters $\alpha_{j,t}$ and $\beta_{j,t}$ in (1) can vary either because there actually exist several generating factors or because the single-factor APT has nonstationary parameters.

Even if nature is single-factor, there are possible nonstationarities which could have a detrimental impact on the quality of estimates derived from fitting (1) to time series of returns, while assuming stationary parameters. First, there is no theoretical reason for the "risk" parameter β_j to be an intertemporal constant. In fact, this parameter is widely acknowledged to depend on factors that are known to change (such as the debt/equity ratio of the firm[4] when asset j is a common stock) and on the absolute risk aversion[5] of investors (which is likely to change with the aggregate level of wealth).[6]

Another source of nonstationarity is in the distribution of ε_j, the disturbance term intended to measure all influences on the return of asset j that are unrelated to the index m.[7] There exists no satisfactory theory about the distribution of ε_j aside from the simple assertion that its mean is zero. In any finite sample, of course, the mean of ε_j can be nonzero insofar as disequilibrium conditions occur temporarily. Also, the higher moments of the distribution of ε_j need not be constant. For example, the variance of ε_j could fluctuate drastically and cause considerable difficulty in econometric estimation of (1).

Finally, of course, nonstationarity can be induced in the model and in estimates of its parameters if the particular form (1) is misspecified. This is the situation when the true state of nature requires several explanatory factors.

There is one very good reason for investigating a single-factor model with data that possibly have been generated by a multifactor process. The reason is simplicity. Just estimating the number of factors is a difficult econometric problem (see Roll and Ross, 1980). Since the single-factor version (1) is so widely applied, a detailed investigation of its parameter stability should be very worthwhile for practical implementation.

Our method of measuring nonstationarity can be adapted easily to a wide variety of problems and improved upon in obvious ways (which we will mention).

The method estimates an explicit time path of a linear model's coefficients and it provides the capability to assess the statistical significance of the nonstationarity. Furthermore, this is done under very weak assumptions concerning the probability distribution that generated the data. No particular generating distribution need be assumed. Such a robust method will clearly not be optimal for every application. An investigator who knows the generating distribution, or is willing to assume that he knows it, will be able to find a specialized technique better adapted to his case; but for those of us with a lower degree of skill or of arrogance, the method to be described here has much virtue.

A. An Outline of What Follows

In the two sections following, we will discuss techniques for avoiding some of the pitfalls caused by nonstationarity. The next section (II) will explain the technique of robust regression; that is, of an improved regression method for a model such as (1) when the disturbances do not have all the standard spherical Gaussian properties so familiar from econometrics texts.

There is no doubt that asset returns, and the estimated residuals from regressions with asset returns, do not have standard properties. Asset returns are *not* stationary Gaussian.[8] There is a controversy about whether the disturbances of (1) are drawn from a stationary non-Gaussian probability distribution, such as a stable distribution, or whether they conform better to a Gaussian process with nonstationary parameters;[9] but whatever the true explanation may be, the observed distributions strongly suggest that an alternative to the classic least-squares method of line fitting may give better results. We will provide evidence that "robust regression" does indeed perform better than ordinary least squares.

In Section III, we describe the second part of a methodology for regression with a nonstationary linear model. It utilizes orthogonal polynomials of time to track the paths taken by the coefficients, the "risk" coefficient β_j, and the intercept parameter α_j. Section IV gives the empirical results for a sample of U.S. assets.

In summary, the basic model (1) will be fit to data while $\alpha_{j,t}$, $\beta_{j,t}$, and the distribution $f_{j,t}(\varepsilon_{j,t})$ of the disturbance term are fit to explicit functions of time. This is a very robust specification and it has the potential to sidestep many of the troubling theoretical and econometric problems previously mentioned. For example, if the "true" state of nature requires another factor (such as in the models of Black, 1972, or of Merton, 1972, or in the more general APT of Ross, 1976) this is econometrically corrected in our specification by allowing the intercept to vary in time. Similarly, if the simple Sharpe-Lintner model, but with nonconstant parameters, is the "true" state of nature, our specification will work. By using model (1) we are not obliged to take sides on the question of which particular theory is "true" because this specification will be an approximation to all of the currently suggested theories (for a particular data sample).

II. ROBUST REGRESSION

The parameters of a linear model are often estimated by the method of ordinary least squares (OLS), which is sensitive to large values of the additive error terms. Various alternative methods have been proposed for obtaining regression estimates which are insensitive to large disturbances and have known sampling properties, at least asymptotically (see Bickel, 1973, and Huber, 1973). The iterative algorithms used in all these studies have the disadvantage that they require a preliminary "reasonable" estimate, usually OLS. Using linear programming to minimize the sum of absolute errors protects against large disturbances, but the only known sampling results were found by artificial data studies (Blattberg and Sargent, 1971). Hinich and Talwar (1975) present an alternative simple two-stage procedure and study its asymptotic and empirical properties when the disturbances are assumed to be independent realizations from a symmetric stable distribution.

Let us present the Hinich-Talwar procedure for estimating the fixed coefficients α and β in the simple model:

$$y_t = \alpha + \beta X_t + \varepsilon_t \qquad t = 1, \ldots, T \tag{2}$$

where ε_t are independent, identically distributed thick-tailed and non-Gaussian errors. In order to obtain the large sample properties of our estimators, the stochastic disturbances are assumed to have a symmetric stable distribution with zero location, scale σ, and characteristic exponent $\gamma < 2$.[10] The assumption of a stable distribution generating process is only incidental. The Hinich-Talwar procedure protects against extreme values of the errors regardless of their distribution if T is large, but some model for the generating process was needed to compute large sample variances and to test the procedure with artificial data. The procedure has been shown to be "robust" against large disturbances (including gross data errors) provided the coefficients α and β are constant over time.

Assuming for convenience that T is even, divide the sample into T/2 nonoverlapping groups of two successive observations (X_t, y_t) and (X_{t+1}, y_{t+1}), $t = 1, 3, \ldots, T - 1$. [Note that in models with k variables $(k > 2)$, the sample would be divided into T/k nonoverlapping groups.] For each t, compute the equation of the line connecting the pair of points; i.e.,

$$\hat{\alpha}_t = \frac{X_{t+1}y_t - X_t y_{t+1}}{X_{t+1} - X_t} \tag{3}$$

and

$$\hat{\beta}_t = \frac{y_{t+1} - y_t}{X_{t+1} - X_t}\,. \tag{4}$$

(And, of course, there would be a vector of k estimated coefficients in the k-variable case.) $\hat{\beta}_1, \ldots, \hat{\beta}_{T-1}$ are independent random variables with the same location β, but with different scales[11] (or dispersion parameters). The scale of β_t is

$$\sigma_t(\beta) = \frac{\sigma 2^{1/\gamma}}{|X_{t+1} - X_t|} \qquad t = 1, 3, \ldots, T - 1.$$

Similarly, $\hat{\alpha}_1, \ldots, \hat{\alpha}_{T-1}$ are independent random variables with location α and scales

$$\sigma_t(\alpha) = \frac{(|X_{t+1}|^\gamma + |X_t|^\gamma)^{1/\gamma}}{|X_{t+1} - X_t|}\,\sigma \quad t = 1, 3, \ldots, T - 1. \tag{5}$$

An initial estimate of β is given by

$$\hat{\beta} = \text{median}\,(\hat{\beta}_1, \ldots, \hat{\beta}_{T-1}),$$

which is a consistent and asymptotically normal estimator: $T^{1/2}\delta_T^{-1}(\hat{\beta} - \beta)$ converges in distribution to N(0,1) as $T \rightarrow \infty$, where

$$\delta_T^2 = \frac{1}{2}\left[\frac{\Gamma(1/\gamma)}{\pi\gamma}\frac{2}{T}\sum_t \sigma_t^{-1}(\beta)\right]^{-2}. \tag{6}$$

If the X_t terms are stochastic, it follows from (6) that for large T, $\hat{\beta}$ is approximately $N(\beta,\delta^2/T)$ where

$$\delta = \frac{\pi\gamma 2^{1/\gamma - 1/2}}{\Gamma(1/\gamma)}\,\sigma\,(E|X_{t+1} - X_t|)^{-1}. \tag{7}$$

The asymptotic efficiency of $\hat{\beta}$ is similar to that for the median estimate of the location given in a random sample from a parent distribution. The efficiency of the first-stage estimator can be increased by using a truncated mean (e.g., the mean of the middle 25% of the ordered $\hat{\beta}_t$ terms, instead of the median $\hat{\beta}_t$). The median was presented here because its asymptotic properties are easier to derive and express. The first-stage estimator of $\hat{\alpha}$ has similar properties.

Once the first-stage estimates of $\hat{\alpha}$ and $\hat{\beta}$ are computed, the residuals,

$$\hat{\varepsilon}_t = y_t - \hat{\alpha} - \hat{\beta}X_t,$$

are ordered and used to compute an estimate of σ,

$$s = \frac{1}{1.654}(\hat{\varepsilon}_{(.72T)} - \hat{\varepsilon}_{(.28T)}),$$

where $\hat{\varepsilon}_{(.72T)}$ and $\hat{\varepsilon}_{(.28T)}$ are the order statistic estimators of the 28th and 72nd percentiles of the distribution of ε. The estimator s has an asymptotic bias of less than .4% for all γ in [1,2]. (See Fama and Roll, 1968, p. 823.)

The sample is then censored by removing all observations (X_t,y_t) corresponding to residuals which are greater in absolute value than some fixed multiple of s. Here, we used 4s as the cutoff. This value was shown in the Monte Carlo experiments of Hinich and Talwar to bring relatively low sampling dispersion to the resulting coefficients over a wide range of distributions of the raw disturbances. The final estimates of α and β are the ordinary least-squares coefficients computed from the remaining observations. They are approximately normally distributed since the errors in the remaining observations have finite variance. As we shall see, this ability to "force" the second-stage least-squares estimates toward normality is of crucial importance in testing for their nonstationarity. Furthermore, the Hinich-Talwar technique will provide this result for any stochastic process of the disturbances ε, not just for stable non-Gaussian disturbances. A nonstationarity in the dispersion of ε, for example, will bring about thicker tails in the sampling distribution of ε. By using this robust procedure, however, the least-squares estimates computed under these conditions will also be forced toward normality.

The procedure may at first seem complex and, of course, the asymptotic distribution theory *is* difficult; but carrying out the required computations will reveal the procedure's basic simplicity. Anyone with access to a regression program can easily implement this robust regression method and virtually guarantee Gaussian disturbances if the number of data observations is sufficient.

III. MODELING NONSTATIONARITY IN THE INTERCEPT, THE SLOPE COEFFICIENT, AND THE MEAN DISTURBANCE

In searching for the best technique to model nonstationarity in the parameters of Eq. (1), there are several useful implications of past empirical work that should be considered. The previous section outlined an approach for accommodating nonstationarity in the dispersion of the disturbance term ε. Here, we wish to present a treatment of nonstationarity in the expected value of ε and in the coefficients.

For common stocks, several studies have documented temporary deviations of $E(\varepsilon)$ from zero. This seems to occur as a result of market disequilibria when new and unanticipated information is received by individual traders. For example, the mean disturbance term is significantly positive in the weeks preceding a stock split.[12] In another example, positive means occur before an announcement of spuriously increased earnings (caused by accounting manipulations) while negative means follow subsequent disclosure of the spurious nature of the increase.[13] Nonzero mean disturbances have been associated empirically with several other occurrences such as secondary offerings[14] and dividend increases,[15] and there are undoubtedly many other circumstances, as yet not documented, which have the same result.[16]

The same arguments can be advanced concerning α_j, the intercept. If the intercept is interpreted as in the Sharpe-Lintner model as a function of the riskless rate of interest, it can vary intertemporally, too. There is probably not as much solid empirical support for its significant variation; but there is certainly no reason to assume constancy in the absence of evidence either way. In several other studies (previously mentioned), the intercept has been presumed to vary with some other stochastic factor. Specification (1) permits the measurement and correction of nonstationarities in both $E(\varepsilon_{jt})$ and in the missing factors (or the riskless return) at the same time. The temporally varying intercept can simply be regarded as a mixture of these influences.

Which time-dependent functions should be chosen to model $\alpha_{j,t}$ and $\beta_{j,t}$? Several alternatives are available: One of the techniques that might have been selected is a modification of the Farley-Hinich (1970) or Farley-Hinich-McGuire (1975) procedures. They allow the coefficients in a regression equation such as (1) to shift *once* during the observed record and the unknown date of the shift can be estimated. This would have been an excellent method if changes in the coefficients of (1) occurred discretely (and only once during the period of observation). There may be cases like this. In fact, such a case would occur for a firm that had floated a new bond issue and thereby changed considerably its leverage ratio during the sample. But there are certainly more complex possibilities and prudence required a more general method.

Another alternative is some type of "random coefficients" technique such as the adaptive regression procedure of Cooley and Prescott (1973). In their method, the coefficients are assumed to vary from one period to the next by following a random walk from their initial positions. Again, the coefficients of some assets may behave this way but others may take on discrete, deterministic changes or fluctuate in a predictable pattern. Besides, the methodology described below is able to track coefficients which actually follow a random sequence. Since it will track other sequences, too, little is lost by its general use. Our idea is to approximate the sequence of each coefficient by a function of time whose parameters can be estimated directly. We decided to use a function of Legendre polynomials for this purpose. The Legendre polynomials are only one of a wide variety of functions that might be used in different applications. We do not imply, by using the Legendre polynomials here, that they are necessarily superior to Chebyshev polynomials, trigonometric polynomials, or many others that could be used to approximate any arbitrary function of time. However, the Legendre polynomials are easy to visualize and they provide an adequate expositional device to illustrate the general principals involved in our technique.

Given any polynomial of time, the time paths of coefficients α_{jt} and β_{jt} can be approximated as

$$\alpha_{j,t} = \sum_{i=0}^{n} a_{ij}P_i(t) \qquad -1 \leqslant t \leqslant 1; \tag{8}$$

$$\beta_{j,t} = \sum_{i=0}^{n} b_{ij}P_i(t) \qquad -1 \leqslant t \leqslant 1; \tag{9}$$

where $P_i(t)$ is the polynomial of order i at time t and a_{ij} and b_{ij} are the coefficients of the ith-order polynomial for the intercept and slope coefficients of (1), respectively. In this case, P is the set of Legendre polynomials and the first five of these are given in Table 1 below and illustrated in Figure 1.

Units of time are chosen so that -1 to $+1$ spans the observed number of natural calendar units.[17] This particular structure is used because the Legendre polynomials are mutually orthogonal on the interval $[-1,1]$.[18] Thus, each polynomial can be introduced as a separate variable in the estimating regression without having to worry about multicollinearity (which would be a serious problem if a nonorthogonal polynomial were used).

The final estimating equation is:

$$R_{j,t} = \hat{a}_{0j} + \hat{a}_{1j}t + \hat{a}_{2j}\frac{3t^2 - 1}{2} + \cdots + \hat{b}_{0j}R_{m,t} + \hat{b}_{1j}tR_{m,t} + \hat{b}_{2j}\frac{3t^2 - 1}{2}R_{m,t} + \cdots + \eta_{j,t} \quad (10)$$

or more simply

$$R_{j,t} = \sum_{i=0}^{n} (\hat{a}_{ij} + \hat{b}_{ij}R_{m,t})\, P_i(t) + \eta_{j,t}. \quad (11)$$

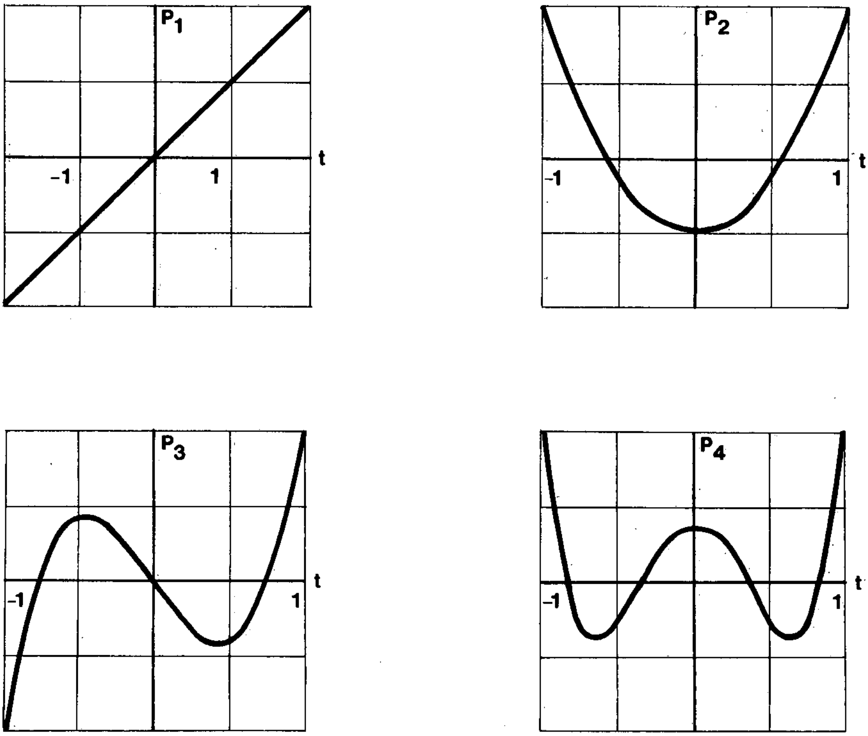
The *averages* of the sequences of the two coefficients $\alpha_{j,t}$ and $\beta_{j,t}$ in the original model (1) are estimated respectively by $\hat{a}_{0j}$ and $\hat{b}_{0j}$ in (11). These would be estimates of $R_F(1 - \beta_j)$ and β_j respectively in model (1) if the simple Sharpe-Lintner model held with a constant R_F and β_j for all t.

The sequence averages $\hat{a}_{0j}$ and $\hat{b}_{0j}$ are not necessarily the best choices to use in prediction. If the true coefficients are actually nonstationary, one should use the most recent estimates of $\hat{b}_{j,t}$ and $\hat{a}_{j,t}$ for purposes of extrapolation. Thus after estimating (11) with a historical record, returns in the next period should be predicted by using

Table 1. Legendre Polynomials, Orthogonal in the Range $[-1, 1]$

$P_0(t) = 1$
$P_1(t) = t$
$P_2(t) = (3t^2 - 1)/2$
$P_3(t) = (5t^3 - 3t)/2$
$P_4(t) = (35t^4 - 30t^2 + 3)/8$

Figure 1. Graphs of Legendre Polynomials.



$$\hat{\beta}_{j,1+\Delta t} = \sum_{i=0}^{n} \hat{b}_{ij} P_i(1 + \Delta t)$$

and by using a similar equation for $\hat{\alpha}_{j,1+\Delta t}$ [where Δt is the prediction interval measured in whatever units of time were used to estimate (11)].

There is one feature of this technique which should be emphasized: the optimal number of high-order polynomials (parameter n above) cannot be determined objectively. A polynomial of higher order than the number of sample points is impossible, of course, but this is a trivial fact since the whole idea is to model nonstationarities with a small number of parameters. In the following results, we have hedged by trying and reporting several alternatives, $n = 0$, 1, and 4. For $n = 0$, Eq. (11) reduces to a particularly simple form of regression model (1) with assumed stationarity in the coefficients (but not necessarily in the dispersion of the disturbances). For $n = 1$, the coefficients can take on a linear trend at most. For $n = 4$, the coefficients can evolve in up to a quartic fashion which will admit a changing trend and even a limited cyclic pattern. In each case, the estimated coefficients of the higher-order polynomials might turn out to be insignificantly different from zero. This would imply stationarity.

Finally, note that a test for the existence of nonstationarity itself during the observed sample period can be made very easily. The composite hypothesis, $b_{1j} = b_{2j} = \cdots = b_{nj} = 0$, using all except the zero-order estimated coefficient in an F test, provides a means of determining the probability that β_j in (1) (the risk coefficient) has actually varied in some way. Similarly, $a_{1j} = a_{2j} = \cdots = 0$ is a hypothesis that neither α_j nor $E(\varepsilon_{jt})$ has changed.

IV. EMPIRICAL RESULTS

A. The Data Sample[19]

The data chosen for illustration here are rates of return (including dividends, if any) for New York Stock Exchange (NYSE)– and American Stock Exchange (AMEX)–listed securities beginning in July 1962. Observations were available daily but were aggregated to weekly intervals.[20] The last observations were for July 1969, and a total of 365 weekly returns were available for a large number of securities. Since there were more than adequate data, we decided to use only a subset of securities and observations, dividing the total period into subperiods and using only those stocks which had a full record of observations. Furthermore, since the computations were done on a commercial computer, we arbitrarily decided to limit the major production run to a fixed number of minutes of CPU time. This had the effect of eliminating American Stock Exchange securities whose names ranked low in alphabetical order since these were positioned last on the magnetic tape.

The final sample consists of 930 securities which have full records during the first 160 weeks (from July 5, 1962, through July 22, 1965). Of the 930 securities, 84 were listed on the American Stock Exchange (and the remainder were on the NYSE). There were some stocks eliminated due to bad spots on the tape.[21]

The reason for choosing 160 weeks was that the robust technique requires the sample size divided by the number of variables to be an integer. Originally, the number of variables was 2, 4, and 10 so that the sample size had to be a multiple of 20. Since 160 was a reasonably large number and a multiple of 20, and since we wanted to reserve at least one-half the available observations (around 365) for prediction and post-sample testing, 160 was chosen. For reasons to be discussed shortly, we later decided to do the robust trimming with the two-variable model only, so we could have chosen a slightly larger sample (and not a multiple of

4 and 10); this seemed hardly worth the extra programming effort, so we did not change the originally chosen sample size. There is no reason to suspect that this rather haphazard technique for choosing the sample had any influence on the results.[22]

The market index used in all cases was the Standard & Poor's (S&P) 500.

B. An Outline of the Results

Since the results comprise a rather large amount of information on several different empirical questions, the following outline will serve as a guide to the most interesting part for each reader:

Section	*Contents*
IV.C	Ordinary least squares (OLS) and Robust (ROB) results for individual stocks for the period July 1962 to July 1965
IV.D	Tests for stationarity in beta coefficients and other parameters for the sample as a whole
IV.E	Analysis of individual securities with many outliers
IV.F	Analysis of individual securities with strongly varying parameters
IV.G	Tests with portfolios; results for the first 160 weeks and postsample refitting
IV.H	Tests of the two-factor model with portfolios

C. Characteristics of Individual Securities, July 1962–July 1965

Table 2 presents cross-sectional statistics (across 930 securities) for regressions calculated with time series for each security. For each time series six different regressions were fit, one set with ordinary least squares (OLS) and another set with the robust technique (ROB). In each set, the three regressions are for (1) the simple index model; (2) the model with linear time trends in the coefficients; and (3) the model with up to a quartic polynomial of time in the coefficients [these three represent model (11) with $n = 0, 1, 4$]. The table presents the calculated coefficients and t statistics for $\hat{a}_{0j}$ and $\hat{b}_{0j}$ from Eq. (11). These coefficients represent the portfolio risk (β_j) and the intercept α_j for the simple (two-variable) model. For the models with nonstationary coefficients, $\hat{a}_{0j}$ and $\hat{b}_{0j}$ are the estimated values of the portfolio risk and the intercept on average over the time series.

For the other coefficients ($\hat{a}_{ij}$ and $\hat{b}_{ij}$ for $i > 0$) only the t statistics are reported in order to save space. Since the entire interest in these coefficients is in their possible deviations from zero and not in their absolute values, reporting the t statistics is sufficient. For each statistic, the cross-sectional arithmetic mean, standard deviation, minimum, 5th percentile, median, 95th percentile, and maximum are given.

Serial correlation in the residuals was insignificant for all models and securities. Over the six different models and the 940 securities, the minimum computed Durbin-Watson statistic was 1.55 and the maximum was 2.76, a range that can be attributed to chance. The mean cross-sectional values of the Durbin-Watson ranged from 2.15 to 2.20 across the six models. This indicates a small and insignificant negative serial dependence in the estimated disturbances (with a first-order correlation of less than .1).

The major differences between ordinary least-squares and robust regression are evident in the table. As mentioned in Sec. II, the robust technique is intended to guarantee that the residuals, and thus the estimated coefficients, follow a Gaussian probability law. The success of this aim is measured by the Studentized ranges calculated from the residuals of each model. For ordinary least squares, the entire cross-sectional distribution of the Studentized range lies to the right of the distribution calculated from the robust residuals.

In the 2-variable model, for example, the 95th sample percentile Studentized ranges are 8.34 for ordinary least squares (OLS) and 5.61 for robust regression (ROB), respectively. Tables of the Studentized range show that the 95th percentile of the null distribution (Gaussian) is near 5.75 for a sample size of 160. This is quite close to the observed value for the robust case but is far below the observed OLS value. For OLS, even the median observed Studentized range is far above the 95th percentile of the null distribution in all these regressions. This indicates, of course, that the observed residuals are quite non-Gaussian when ordinary least squares is employed with these data. Thus the computed OLS t statistics are not necessarily reliable measures of significance. With the robust technique, however, the residuals do not violate the assumption of normality. Thus tests of significance based on the normality assumption have face validity. This is very important in testing for the presence of nonstationary coefficients, as we shall see in the next subsection.

One other fact about the Studentized range should be noted: Its cross-sectional standard deviation is three times larger for OLS than for ROB. This is no doubt due to some extremely large values (the observed maximum for OLS is 12.7), which indicate gross departures from normality for some securities. Whatever the source of these departures, be it non-Gaussian stable probability laws, data errors, or nonstationarities in the parameters, the results make a strong case in support of the use of the robust technique, particularly for securities with very large observed values of the Studentized range (and correspondingly large potential errors in estimated parameters).

The explanatory power of the regression seems also to be improved by the robust technique. This is shown by the observed cross-sectional distribution of R^2, which is shifted to the right from the OLS distribution.

Table 2. Index Models with Possibly Nonstationary Parameters for New York and American Stock Exchange Securities,[a] July 1962–July 1965[b]

Estimated Parameter	*Refers to Coefficient*[c]	*Regression Technique*[d]	*Mean*	*Standard Deviation*	*Minimum*	*5th Percentile*	*Median*	*95th Percentile*	*Maximum*
			2-Variable Model[e] (Assumes Stationary Parameters)						
Intercept (%/Annum)	$\hat{a}_0$	OLS	-.151	15.3	-71.0	25.4	.390	24.6	57.1
		ROB	-8.94	19.7	-98.0	-47.6	-5.18	17.7	55.9
t (Intercept)	$\hat{a}_0$	OLS	.0371	.950	-3.02	-1.50	.0338	1.56	2.90
		ROB	-.574	1.30	-5.02	-2.84	-.469	1.39	3.39
Slope	$\hat{b}_0$	OLS	1.05	.442	-1.04	.415	1.04	1.79	2.59
		ROB	.955	.448	-.552	.329	.969	1.78	2.76
t (Slope)	$\hat{b}_0$	OLS	5.17	2.14	-2.57	1.85	5.06	8.78	14.7
		ROB	5.53	2.29	-2.75	2.00	5.44	9.43	15.8
Adjusted R^2		OLS	.143	.0959	-.0127	.0109	.129	.320	.574
		ROB	.163	.105	-.0136	.0132	.151	.357	.614
Studentized Range		OLS	6.40	1.11	4.44	5.04	6.21	8.34	12.7
		ROB	5.09	.303	3.89	4.62	5.07	5.61	6.21
			Mean Coefficients, 4-Variable Model[e] (Linear Trend)						
Intercept (%/Annum)	$\hat{a}_0$	OLS	-.333	15.5	-72.4	-26.8	.152	24.5	58.4
		ROB	-9.17	19.9	-97.8	-48.5	-5.43	17.5	55.9
Slope	$\hat{b}_0$	OLS	1.04	.450	-1.08	.394	1.03	1.82	2.77
		ROB	.995	.456	-.617	.334	.968	1.79	2.70

					Statistics, 4-Variable Model[e]				
t (Intercept)	$\hat{a}_0$	OLS	.0255	.961	−3.07	−1.55	.0152	1.58	2.97
		ROB	−.589	1.32	−5.00	−2.92	−.474	1.41	3.37
t (Intercept) Trend	$\hat{a}_1$	OLS	.204	.866	−2.85	−1.22	.234	1.50	2.83
		ROB	.243	.938	−.284	−1.33	.276	1.78	3.02
t (Slope)	$\hat{b}_0$	OLS	4.82	2.03	−2.75	1.64	4.70	8.24	13.6
		ROB	5.20	2.19	−2.89	1.85	5.12	8.92	14.8
t (Slope Trend)	$\hat{b}_1$	OLS	−.192	1.26	−4.38	−2.22	−.173	1.83	3.57
		ROB	−.146	1.25	−4.25	−2.24	−.100	1.88	4.36
Adjusted R^2		OLS	.145	.0972	−.0241	.00926	.130	.326	.579
		ROB	.166	.106	−.0229	.0124	.154	.361	.619
Studentized Range		OLS	6.40	1.11	4.44	5.04	6.21	8.34	12.7
		ROB	5.12	.304	3.85	4.66	5.11	5.66	6.24
					Mean Coefficients, 10-Variable Model[e]				
Intercept (%/Annum)	$\hat{a}_0$	OLS	.581	15.5	−74.7	−25.7	1.25	24.4	56.1
		ROB	−8.54	19.9	−97.9	−48.0	−4.82	17.1	54.7
Slope	$\hat{b}_0$	OLS	1.01	.471	−.186	.327	.976	1.83	2.70
		ROB	.969	.469	−.929	.292	.925	1.82	2.72
					Statistics, 10-Variable Model[e]				
t (Intercept)	$\hat{a}_0$	OLS	.0950	.949	−3.12	−1.43	.100	1.56	2.99
		ROB	−.521	1.30	−4.79	−2.77	−.431	1.42	3.24

Table 2. Index Models with Possibly Nonstationary Parameters for New York and American Stock Exchange Securities,[a] July 1962–July 1965[b]

Estimated Parameter	*Refers to Coefficient*[c]	*Regression Technique*[d]	*Mean*	*Standard Deviation*	*Minimum*	*5th Percentile*	*Median*	*95th Percentile*	*Maximum*
				Statistics, 10-Variable Model[e]					
t (Intercept) Trend)	$\hat{a}_1$	OLS	.239	.886	− 2.81	− 1.24	.245	1.58	2.91
		ROB	.277	.946	− 2.81	− 1.28	.315	1.76	3.26
t (Intercept) Quadratic)	$\hat{a}_2$	OLS	− .0293	.833	− 2.73	− 1.32	− .0572	1.34	3.06
		ROB	.0220	.926	− 2.45	− 1.42	− .00402	1.54	3.42
t (Intercept Cubic)	$\hat{a}_3$	OLS	− .213	.844	− 2.81	− 1.58	− .221	1.17	2.58
		ROB	− .180	.872	− 2.77	− 1.54	− .178	1.20	2.25
t (Intercept Quartic)	$\hat{a}_4$	OLS	− .200	.774	− 2.86	− 1.52	− .179	1.06	2.97
		ROB	− .201	.822	− 2.61	− 1.58	− .185	1.16	2.62
t (Slope)	$\hat{b}_0$	OLS	4.20	1.90	− 2.65	1.34	4.08	7.35	12.5
		ROB	4.57	2.04	− 1.99	1.41	4.50	8.06	13.7
t (Slope) Trend)	$\hat{b}_1$	OLS	− .114	1.15	− 4.16	− 2.07	− .0999	1.74	3.44
		ROB	− .0859	1.18	− 4.19	− 2.00	− .0805	1.72	3.70
t (Slope Quadratic)	$\hat{b}_2$	OLS	.374	1.10	− 3.31	− 1.48	.363	2.12	4.08
		ROB	.331	1.10	− 3.31	− 1.45	.351	2.11	3.58
t (Slope Cubic)	$\hat{b}_3$	OLS	.0416	1.01	− 3.37	− 1.60	.0642	1.71	3.84
		ROB	.0609	1.03	− 3.13	− 1.54	.0671	1.76	3.86
t (Slope Quartic)	$\hat{b}_4$	OLS	− .169	1.09	− 4.08	− 1.97	− .162	1.67	3.06
		ROB	− .148	1.07	− 3.67	− 1.86	− .167	1.65	2.93

Adjusted R^2	OLS	.144	.100	−.0517	.00111	.132	.327	.588
	ROB	.166	.108	−.0541	.00842	.153	.366	.616
Studentized Range	OLS	6.39	1.10	4.50	5.04	6.20	8.33	12.7
	ROB	5.18	.324	4.32	4.67	5.17	5.72	6.52

[a] 930 individual securities.
[b] 160 weekly observations.
[c] This refers to the coefficient from the fitted equation (11).
[d] OLS = ordinary least squares; ROB = Hinich-Talwar robust regression.
[e] The 2-, 4-, and 10-variable models are Eq. (11) with n chosen as 0, 1, and 4, respectively.

Of course, this is hardly surprising since the robust technique has the effect of throwing out observations that do not conform well to the estimated regression hyperplane. One should be very careful to note that this supposedly higher explanatory power is really just a bias in the R^2 computed by the robust technique. If the discarded observation had been added back to the computation of R^2 (using estimated residuals for these observations that were calculated with the robust coefficients), the result would be a value for R^2 much closer to the OLS value. For the same reason, the predictive ability of the ROB equations will not be better than the OLS equations to the extent of the "improvement" in R^2. This is because the same processes which generated outliers during the fitted sample period can generate outliers in the postsample period. If there is any improvement in predictive ability by going from OLS to ROB, it will only be because the ROB coefficients have less estimation error.

We began the study by using the robust method to compute outliers for each model (2, 4, and 10 variables) separately. Unfortunately, this is a very treacherous procedure for the 4- and 10-variable models, as we soon discovered. The reason stems from the orthogonal polynomials employed as regressors in these models. Recall that the robust technique first partitions the sample into N/k subsamples, where N is the total sample size and k is the number of variables. A regression is then computed for each subsample and the medians of these coefficients are taken as the first-pass estimates. Unfortunately, even though the polynomials are orthogonal over the full time span of 160 observations, they are far from orthogonal within each subperiod.

This is quite apparent from a glance at the graphs of the Legendre polynomials (Figure 1). If the subsamples for the robust method are chosen by a natural partition along the time axis, the intercorrelation among the polynomials within each subsample will be quite high. For the 10-variable case, the axis is partitioned into 160/10 = 16 equal subsections. A 1/16th subsection of the time axis, chosen anywhere at all, will result in extremely high interdependencies among the polynomials. The 4-variable case is not as troubling because only the first-order polynomial and the first-order polynomial multiplied by the market return are used. However, the subsections are smaller, each being 1/40th of the total sample. Thus, some nearly singular matrices have a high probability of occurrence, simply because four successive market returns of approximately the same magnitude are rather likely to occur at least once.

We tried to remedy this induced multicollinearity first by choosing subsamples by some method other than the natural ordering along the time axis. For example, we tried using observations numbered 1, 11, 21, ..., 151 for the first subsample, 2, 12, 22, ..., for the second subsample, etc. in the 10-variable case and 1, 5, 9, etc. for the 4-variable regressions.

This alleviated considerably the multicollinearity problem for the 10-variable case but it did not help the 4-variable case at all, and there were still some subsamples in both cases whose moment matrices had extremely low determinants.

Finally, we decided to do the first part of the robust technique, the determination of outlier observations, with the 2-variable model alone. Then the second pass, which consists of computating ordinary least squares on nonexcluded observations, was done for the 2-, 4-, and 10-variable model but using the observations determined by the 2-variable first pass. Admittedly, this is a somewhat arbitrary procedure and one would expect a priori that it would understate the true worth of the robust technique relative to OLS for the 4- and 10-variable models. However, the special nature of the independent variables in these models made it necessary and we can do no better than to keep it in mind when interpreting the results.

D. Testing for the Existence of the Nonstationary Parameters A First Pass

One of the basic attractions of our method is its capability to provide a statistical test for the existence of nonstationarity. We illustrate this by tests of stationarity of the coefficients of (1). In particular, the stability of "risk" parameter [β_j in (1)] over time is a matter of great importance to portfolio managers and other investors. They would like to be able to use historically estimated risk parameters for current portfolio selection and would also like to be able to extrapolate them safely. In addition, the intercept term α_j is sometimes related to an "extraordinary" return for a stock, above and beyond the normal compensation for its risk. Thus, it too is important in some contexts. We will report evidence on the stationarity of both parameters.

The central problem in testing for the overall existence of nonstationarity is the a priori lack of knowledge of the direction of change. During the sample period, we might be lucky and observe risk coefficients on all securities drifting together. Certain reasons for drift, such as changing attitudes toward risk on the part of all investors, might reasonably be supposed to cause such a general and concurrent movement. If this did occur, we would observe that some or all of the coefficients associated with polynomials of time in the estimating equation (11) were significantly nonzero on average.

Other reasons for nonstationarity, such as changes in individual firms' capital structures, would produce nonconcurrent movements in risk coefficients. Some coefficients might increase and others decrease, depending on the specific circumstances of each firm. Since the average value of the risk coefficient must be very close to unity, these nonconcurrent nonstationarities would not cause the mean cross-sectional values of the polynomial coefficients of (11) to deviate from zero. They would, however, cause the estimated polynomial coefficients to follow a different cross-sectional *distribution* than that expected under the null hypothesis of nonstationarity. The distribution might very well be located at zero but the cross-sectional dispersion would be larger than the anticipated null dispersion.

This implies that tests for nonconcurrent nonstationarity are highly dependent on a knowledge of the distribution of the estimates from (11). We have already seen, however, that ordinary least squares applied to our data produces regressions with non-Gaussian residuals (and thus with "t ratios" that do not necessarily follow Student's distributions). In contrast, the robust regression technique guarantees asymptotic normality. Furthermore, Studentized ranges reported in Table 2 support the contention that our chosen sample size is sufficient to make valid an assumption of Gaussian disturbances from the robust models. Therefore, we will use the robust results in testing for nonstationarity since we can be reasonably confident about the sampling distribution of the t ratios for this technique.

Under the null hypothesis of totally stationary coefficients, the higher-order robust-estimated coefficients of (11) ($\hat{a}_{1j}$, $\hat{a}_{2j}$, ..., $\hat{b}_{1j}$, $\hat{b}_{2j}$, ...) are normally distributed with mean zero and standard deviations estimated by their computed standard errors. The t ratios are distributed according to the Student law with about 146 and 140 degrees of freedom for the 4- and 10-variable models, respectively. (Degrees of freedom differ across securities according to the number of rejected outliers.) This Student law is very close, but not exactly equal, to the standardarized Gaussian law. For df = 120, the 90% interfractile range is 3.316 for Student while it is 3.290 for standardarized Gaussian. The difference is smaller for all but one of the 930 securities since the minimum sample size is 128 and the next lowest sample size is 134. Therefore, using the Gaussian approximation will result in only a trivial numerical bias.

The 90% range is itself asymptotically normal with standard deviation approximately equal to

$$\hat{\sigma}_{R.9} = \frac{1}{f_{.05}}\left[\frac{2(.045)}{N}\right]^{1/2} = .234$$

where $f_{.05}$ is the standard Gaussian ordinate at the 5th percentile and N is the sample size (which we have taken equal to 155 in this calculation).[23]

Confidence regions for the ranges of estimated coefficients can be constructed with these numbers. For example, the range 3.290 ±

(.234)(1.96), or [2.83,3.75] is the 95% acceptance region of the observed 90-percentile range for the null hypothesis of stationarity. If the observed range falls within this interval, the null hypothesis cannot be rejected. Table 3 presents test statistics associated with this idea. The table gives computed values of a "standardized" 90-percentile range, z, defined as

$$z \equiv \frac{\hat{R}_{.9} - 3.290}{.234}$$

where $\hat{R}_{.9}$ is the estimated 90-percentile range (computed as the difference between the 95th and 5th cross-sectional percentiles of the t ratios of Table 2). Values of z are starred if they imply an observed range outside the 95% acceptance interval and thereby reject the hypothesis of stationarity.

The table also presents tests labelled $\bar{t}/(\hat{\sigma}_t/\sqrt{N})$ for the significance of the cross sectional mean t ratio and $\bar{\hat{c}}/(\hat{\sigma}_c/\sqrt{N})$ for the significance of the cross-sectional mean coefficient. The symbol t denotes the arithmetic mean t ratio in Table 2 and $\hat{\sigma}_t$ denotes the cross-section standard deviation of the observed t ratios. $\bar{\hat{c}}$ is the arithmetic cross-sectional mean of the

Table 3. Tests for the Presence of Nonstationary Parameters among 930 NYSE and AMEX Securities, 1962–1965 (Robust Regression)

Estimated Parameter	*No. of Variables in Model*	*Standardardized .9 Range (z)*	$\frac{\bar{t}}{\hat{\sigma}_t/\sqrt{N}}$	$\frac{\bar{\hat{c}}}{\hat{\sigma}_c/\sqrt{N}}$
Intercept Trend, $\hat{a}_1$	4	− .769	7.90*	9.09*
Slope Trend, $\hat{b}_1$	4	3.54*	−3.56*	−2.65*
Intercept Trend, $\hat{a}_1$	10	−1.07	8.93*	9.74*
Intercept Quadratic, $\hat{a}_2$	10	−1.41	.725	1.91
Intercept Cubic, $\hat{a}_3$	10	−2.35*	−6.29*	−5.63*
Intercept Quartic, $\hat{a}_4$	10	−2.35*	−7.45*	−5.69*
Slope Trend, $\hat{b}_1$	10	1.83	−2.22*	−1.47
Slope Quadratic, $\hat{b}_2$	10	1.15	9.18*	7.94*
Slope Cubic, $\hat{b}_3$	10	.0427	1.80	1.49
Slope Quartic, $\hat{b}_4$	10	.940	−4.22*	−3.19*

* Rejection of the hypothesis of stationarity at the 95% level of significance.

estimated coefficient and $\hat{\sigma}_c$ is the cross-sectional standard deviation. Values significantly different from zero at the 95% level are starred.

In comparing these tests with the range test, one can think of the range as measuring nonconcurrent changes (across securities) in the model's parameters while the mean t ratio and mean coefficient measure concurrent changes.

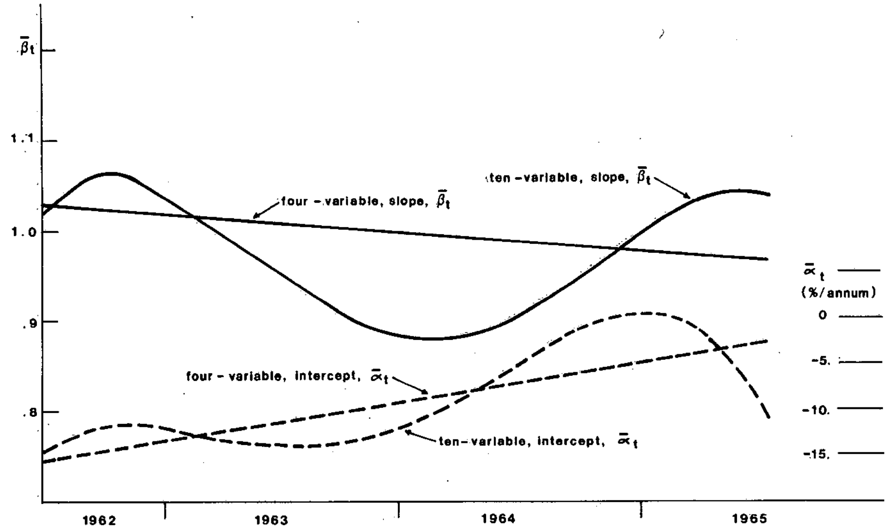
In both the 4-variable and 10-variable models, there seems to be evidence of significant nonstationarity of both types. For example, the "slope trend" [associated with the first-order polynomial coefficient $\hat{b}_{1j}$ in (11)] has a t ratio which is, on average, significantly negative. The 10-variable results indicate that the average drift in the risk coefficient was not a simple linear function of time (as it must be regarded by the 4-variable model which can accommodate only the first-order polynomial). Both the quadratic and the quartic polynomials associated with slope changes had significant mean t values in the 10-variable model. In fact, the cross-sectional mean slope trend is only marginally significant in the 10-variable model, which means that most of its variation was nonlinear.

The mean intercept also seems to have changed significantly (and nonlinearly) during the sample period. This is of course not too surprising since the intercept should be a negative linear function of the slope according to some theories. Indeed, the first-order polynomial coefficients are in agreement, since they are positive for the intercept change and negative for the slope change. They are more significant for the intercept than for the slope, however, and this could indicate a nonstationary riskless rate of interest or the movement of some other omitted factor.

The time paths implied by these mean coefficients are given in Figure 2. This plot simply assumes that mean estimated coefficients of the higher-order polynomials accurately depict the actual paths of the slope and intercept during the sample period. We must enphasize that this graph is biased toward showing no movement. If the index used as the explanatory variable had been composed of the stocks in the sample, the mean cross-sectional $\beta_{j,t}$ would have been unity for every period, by construction. Thus it could not have been nonstationary. Figure 2 shows movement only because the index was not composed of the sample of securities weighted in the same proportion.

The tests for nonconcurrent changes also find a few significant values. The slope trend in the 4-variable model, for example, has a larger 90% range than one should have expected under the hypothesis that all coefficients are stationary. This agrees with the first two polynomials from the 10-variable model. (They are significant together but not individually.) Evidently, in addition to many securities displaying concurrent drift,

Figure 2. Time Paths of Mean Coefficients, 930 Securities, 1962–1965.



there were also some which drifted significantly in a direction opposite to the average. This implies that the *concurrent* drift was actually more substantial than the average coefficient indicates since the average coefficient contains some securities which moved in opposition.

There is unfortunately a statistical problem with these tests: they assume that the cross-sectional distribution of estimated coefficients constitutes a random sample. If there is interdependence among the estimates, both tests will be biased (but in different directions). For example, if there is positive dependence among the estimates, the computed standard error of the mean t value ($\hat{\sigma}_t/\sqrt{N}$) will be lower than the standard error that would have been observed in a genuine random sample. This will overstate the significance of the mean t value for all securities (and thus will make the concurrent drift appear more significant than it really is). On the other hand, the standard error of the sample .9 range does not depend on the observations. (It is computed from the known distribution of a random sample of order statistics from a Gaussian law.) Therefore, positive dependence among the estimates would cause this test statistic to *understate* the true significance of nonconcurrent movements (because the observed range would be smaller than a range obtained from a random sample). The actual interdependence, if any, would depend on the cross-sectional dependence in the disturbances. King (1966) found a very small degree of positive cross-sectional dependence in the disturbances from a version of the 2-variable model but even a small degree of covariation can cause a large bias in the estimated significance of the mean coefficients or the mean t ratios. However, since the presence of interdependence would bias the range test in an opposite direction, and since both types of tests indicate some significant nonstationarities in the coefficients, we are safe in concluding that *some* kind of nonstationarity was present during the sample period, even though we should be hesitant to state that it was concurrent rather than nonconcurrent. In Section IV.G this question will be reexamined with estimates obtained from portfolio returns, which will be used to obtain an estimate of the cross-sectional interdependence.

E. Detecting Unusual Observations

Another potentially useful capability of the technique is the identification of unusual cases. When the same model is applied to a large number of cases, an automatic method for detecting departures from the norm would be highly valuable. For example, a portfolio manager could automatically detect alterations in the return process that would prompt a more extensive investigation via fundamental analysis of the company.

Our technique provides two indications of such "unusual" circumstances. The first, to be derived in this section, relies on observations detected by the robust procedure as lying outside the normal range. The second, to be discussed in the subsequent section, involves unusual movements in the estimated coefficients of the linear model.

Again reverting to our example model (1) with stock return data, the cross-sectional distribution of the number of nonexcluded observations from the first-pass robust regression is given in Table 4.

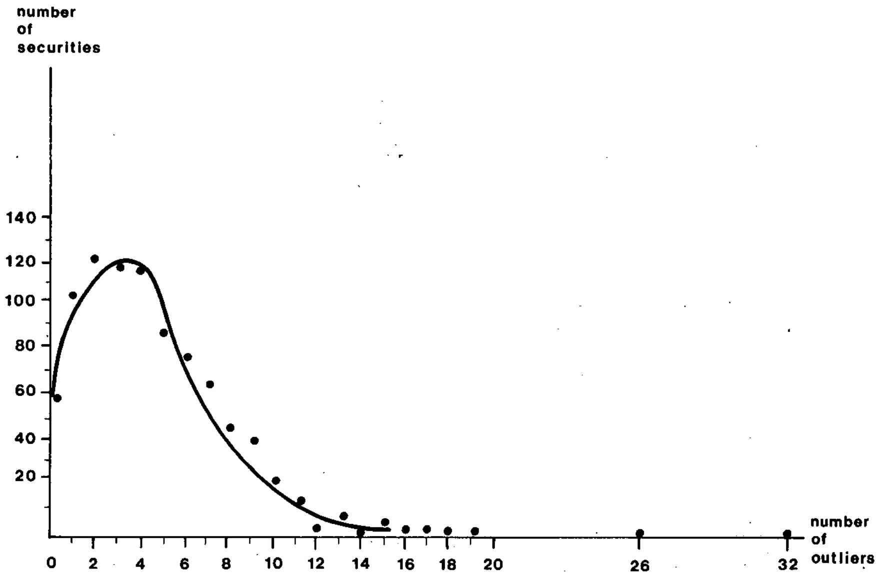
Figures 3 and 4 provide further information about the effect of robust regression. Figure 3 gives the cross-sectional frequency distribution of the number of outliers (which is 160 minus the number of included observations). This shows that 58 of the 930 securities had no excluded observations. For these securities, of course, the OLS and robust estimates were identical. The most frequent number of outliers was in the range 2 to 4, which is a very small percentage of the sample. There were

Table 4. The Number of Nonexcluded Observations[a] Using the Robust Method on 930 NYSE and AMEX Securities, July 1963–July 1965

Mean	*Standard Deviation*	*Minimum*	*5th Percentile*	*Median*	*95th Percentile*	*Maximum*
155.	3.50	128.	149.	156.	160.	160.

[a] Maximum possible is 160.

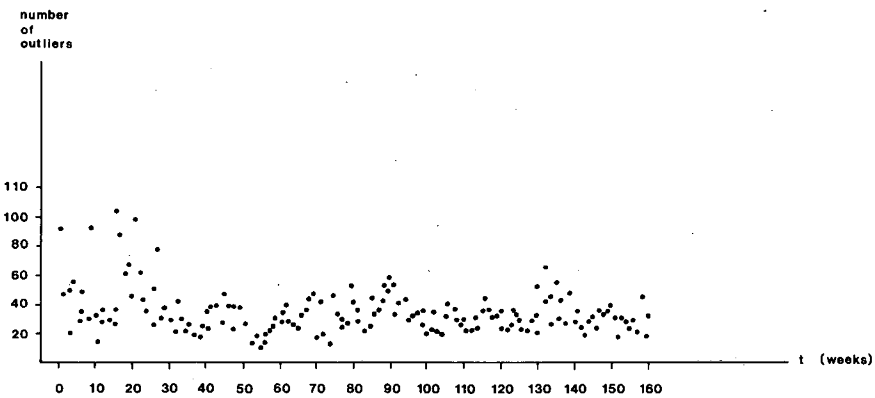
Figure 3. Frequency Distribution of Number of Securities vs. Number of Outliers: 930 NYSE and AMEX securities, 1962–1965.



some securities, however, with a significant proportion of excluded observations. They will be the subject of further analysis in Section IV.E.

Figure 4 presents the number of outliers as a function of time. The interest in this chart derives from possible nonstationarities in the dispersions of individual security disturbance terms. If such nonstationar-

Figure 4. Number of Outliers by Week from 930 NYSE and AMEX stocks, 1962–1965.



ities are interrelated across securities, we should observe a pattern in the time path of the number of outliers. Indeed, Fig. 4 shows that the 5 weeks with the most outliers occurred before week 30 (which ended January 24, 1963). There is also some serial dependence over time in the number of outliers. This would be consistent with, for example, nonstationary but slowly varying dispersions of disturbances for all securities.

To check whether part of this is due to unusual movements in the S&P 500 index, caused by abnormal changes in one or more of its component securities, we calculated the following relation between the natural logarithm[24] of the number of outliers, N, and the current and lagged absolute values of the index return, R_m (the absolute value of the market index return is a proxy for its dispersion):

$$\log_e(N_t) = \underset{(50.1)}{3.05} + \underset{(1.39)}{5.81}\,|R_{m,t}| + \underset{(2.20)}{9.14}\,|R_{m,t-1}| \qquad t = 2, \ldots, 160,$$

where Durbin-Watson (DW) = 1.15; Studentized range (SR) = 5.79; the numbers in parentheses are t ratios.

Because of the low Durbin-Watson, the regression is misspecified. As a rough remedy, we added a lagged dependent variable as a regressor. This resulted in

$$\log_e(N_t) = \underset{(7.82)}{1.74} + \underset{(1.05)}{3.96}\,|R_{m,t}| + \underset{(1.49)}{5.66}\,|R_{m,t-1}| + \underset{(6.08)}{.425}\,\log_e(N_{t-1})\big|_{t=2,\ldots,160};$$

$$DW = 2.13,\ SR = 5.79.$$

The significance of the lagged value of the number of outliers justifies our previous statement about serial dependence. This regression shows also that the influence of the market on the number of outliers is not very strong, which implies in turn that most of the outliers are due to events specific to individual securities or specific to groups of securities in the sample and are not due to unusual changes in the general market index. Note well that the average value of $R_{m,t}$ is about .008 while the average value of $\log_e(N_t)$ is about 1.6. Thus, even though the coefficients of $|R_{m,t}|$ and $|R_{m,t-1}|$ are larger than the coefficient of the lagged value of the number of outliers, their average affect on the dependent variable is only about 1/20th as great.

Thus, we have demonstrated that the robust technique eliminates, by a systematic and prespecified method, sample observations which are not congruent with the majority of observations. The actual number of "outliers" varies across securities. To a certain extent, part of this variation must be ascribed to chance, since random differences among the 930 securities are bound to occur. It is also true, however, that securities with a large proportion of rejected observations are more likely than other securities to have experienced unusual and unique events with greater frequency during the sample period. In an effort to identify such unique circumstances, we have separated out for detailed analysis all securities with more than 10 outliers (out of 160 observations) during the sample period. There were 48 securities in this class, or slightly more than 5% of all securities. They are given by name in Table 5, along with the numbers of outliers observed for each one. (In any other application, of course, the percentage of cases to be singled out for further analysis could be different. Here, we are merely illustrating the semiautomatic nature of the culling process.)

The subsample consisting of these 48 high-outlier stocks differs in some important respects from the total sample. Some differences could have been foreseeable. For example, the mean Studentized ranges of

Table 5. Securities with Many Unusual Observations, NYSE and AMEX, July 1962–July 1965

Name	*No. of Outliers*	*Name*	*No. of Outliers*
Acme Precision Products*	17	General Portland Cement	13
Allied Control Co.*	13	Ginn & Co.	15
Amerace Corp.	12	Gordon Jewelry*	14
American Bakeries	11	Great Basins Petroleum*	17
Ancorp Nat'l Services	13	Great Northern Iron Ore	15
Argus, Inc.*	18	Hartfield Zodys, Inc.*	11
Atlas Corp	19	Indiana General Corp.	14
Baltimore & Ohio RR	14	Libby, McNeil & Libby	16
Bausch & Lomb	11	Meritt-Chapman	32
Beatrice Foods	11	Minn Enterprises	14
Bell Intercontinental	15	Monon RR (B)	15
Briggs Mfg.	13	J. J. Newberry	11
City Stores	17	Oklahoma Gas and Elec	13
Colt Industries	14	Pacific Tin	15
Continental Materials*	11	Random House, Inc.	11
Countrywide Realty*	15	Rapid American	18
Dorr-Oliver, Inc.	11	Reliable Stores	15
Electronic Commun.*	11	Standard Oil of Ohio	13
El Paso Natural Gas	11	Stone & Webster	12
Esquire, Inc.	11	Texas Pacific Land	11
Family Finance	11	Tractor Supply Co.	12
Federal-Mogul	12	Transwestern Pipe Line	13
Fischbach & Moore	12	Udylite Corp.	12
Forest City Entreprises*	11	Welbilt Corp.	26

* American Stock Exchange–listed security.

residuals computed by ordinary least squares (OLS) are 7.22, 7.22, and 7.21 for the 2-, 4-, and 10-variable models respectively. This is .82 higher than the OLS mean Studentized ranges for all securities (see Table 2), which implies that disturbances for these high-outlier securities deviate further from Gaussian than do all securities. After application of the robust (ROB) method, the mean Studentized ranges across the 48 securities drop to 5.08, 5.08, and 5.16. These are very close and even slightly less than the mean Studentized ranges computed with the robust method on all 930 securities. Thus, even though the disturbances seem to have been very non-Gaussian over the full sample record and more non-Gaussian for the high-outlier group than for all securities, the rejection of a relatively large number of observations seems to have brought the desired result of Gaussian-distributed estimates. The high-outlier subsample also differs from the sample of all securities in ways that could not have been foreseen. The average proportion of explained variation (R^2) is much lower for these securities, being less than .07 for OLS and about .08 for ROB. This is only half the proportion of explained variation for all securities.

Table 6 presents evidence on still another significant difference, the absolute magnitude of the mean portfolio risk coefficient (β). For this 48-security subsample, the mean risk coefficients are significantly lower than for all securities (cf. Table 2). Furthermore, the estimates obtained by the robust method are significantly less in absolute magnitude than those obtained from OLS. This is not matched, incidentally, by lower t values for the robust coefficients.[25] In fact, the robust estimates are slightly more significantly different from zero than the OLS estimates although both are less significant than the average coefficients for all securities. This is a somewhat puzzling result. For a group of securities with relatively large numbers of unusual observations, we might have guessed that ROB would have provided more *accurate* estimates of risk coefficients than OLS.[26] But we would not have expected to find a significant difference in their absolute values. Furthermore, one can easily see from Table 6 that the difference is not caused by just a few unusual estimates out of the 48. Every fractile (except the minimum) is lower for ROB than for OLS.

F. Strongly Varying Parameters

The second type of individual case that could be regarded as "unusual" involves highly significant changes in the coefficients. Table 7 lists securities for which estimated higher-order polynomial β coefficients had t ratios in either the OLS or the robust regressions greater than 3.5 during the 160-week sample period. These securities displayed more significant nonstationarity than all others.

Table 6. Estimated Risk Coefficients[a] for the 48 High-outlier Securities, NYSE and AMEX, 1962–1965

Model (No. of Variables)	*Regression Technique*	*Mean*	*Standard Deviation*	*Minimum*	*5th Percentile*	*Median*	*95th Percentile*	*Maximum*
2	OLS	.823	.438	−1.04	.0918	.904	1.33	1.64
	ROB	.596	.345	− .502	−.219	.627	1.00	1.43
4	OLS	.808	.451	−1.08	.0598	.819	1.33	1.58
	ROB	.603	.344	− .501	−.0369	.606	1.05	1.35
10	OLS	.743	.540	−1.86	.0478	.772	1.45	1.54
	ROB	.564	.365	− .929	−.00750	.573	1.00	1.36

[a] I.e., Estimated coefficients $\hat{b}_{0,j}$ of the zero-order polynomial in model (11).

Table 7. NYSE-listed Securities with Strongly Varying Parameters, 1962–1965

Security	Estimated Change in Beta (Robust)		Change in Market D/E Ratio, 1962–1965	Linear Significance $\lvert t\rvert > 2.0$	
	4-Variable	10-Variable		OLS	ROB
American Commercial Lines	2.12	2.16	.171	√	√
American South African Investment	− .413	2.64	0.		
CCI Marquardt	−5.11	−5.81	− .0935	√	√
Crescent Corp.	− .785	1.70	− .823		
Dan River Mills	1.78	2.00	− .0947	√	√
Dorr-Oliver	− .0736	.556	− .0747		
Florida Power	−1.36	−1.33	− .038	√	√
W. R. Grace	−1.74	−1.18	− .173	√	√
Madison Fund	− .862	− .664	0.	√	√
N. Y. Central	2.56	2.25	−2.62	√	√
Niagara-Mohawk Power	− .0574	1.07	− .217		
Pacific Tel. & Tel.	− .853	− .896	.088	√	√
Peoples Gas, Light & Coke	−1.04	− .877	− .036	√	√
Proctor & Gamble	−1.24	−1.13	− .0019	√	√
Richardson-Merrill	− .353	− .973	0.		
Tractor Supply Co.	− .872	− .270	.0776	√	
Univ. Leaf Tobacco	−1.90	−2.10	− .775	√	√
Zapata Norness Inc.	3.42	2.04	− .166	√	√

Some of these securities may have entered this group by chance. Whenever the extreme values of a sample distribution are isolated, some members of the group will be there simply from random variation. We know, however, from the general tests for the existence of nonstationarity reported in Sec. IV.D, that some securities really did have changing coefficients.

There are two possible ways to determine whether a particular security had genuine nonstationarity. First, if a polynomial coefficient is significant in the OLS and not the robust regression, the probability is greater that the measured nonstationarity has arisen by chance. This conclusion is based on the fact that the OLS t ratios do not necessarily conform to Student's distribution and are not, therefore, necessarily subject to a known acceptance-rejection region. In addition, the robust coefficients are more reliable because data errors, bad spots on the tape, or grossly abnormal observations are censored by the robust calculation.

The second possibility for discriminating genuine from false cases is a detailed analysis of peculiar events for each security. As we mentioned early in the chapter, there are known influences on the beta coefficient and these can be examined directly for change. If, for example, a major issue of new debt brought about significantly greater capital leverage during the period of observation, we should find this corresponding to a positive change in the beta.

The numbers in Table 7 give the total estimated change in beta, the portfolio risk coefficient from the first date to the last date in the sample period (July 12, 1962–July 29, 1965). For the 4-variable model, which permits a linear change, this is simply twice the estimated slope of the change with respect to time.[27] For the 10-variable model, it is the difference between the estimated time track of the coefficient at the last and first dates. Also noted in the table are significant (with $|t| > 2.0$), linear polynomial coefficients ($\hat{b}_{1j}$) for ordinary least-squares and robust regression.[28] This table is our attempt to present compactly the salient facts from the 6 regressions and 32 coefficients estimated for these 18 securities. The facts are these:

1. For 12 of the 18 securities, there is evidence of a strong *linear* change in the beta coefficient. These securities have linear coefficient absolute t ratios in excess of 2.0 for both OLS and robust. In most of these cases, the absolute coefficient changes over the sample period, estimated with the 4- and 10-variable models, agree quite closely. Nine of 12 differ by less than 15%.
2. One security, Tractor Supply Company, has significant OLS linear coefficients but the corresponding robust coefficients are not significant at all.[29] The estimated negative change in the coefficient is 3 times larger for the 4-variable than for the 10-variable model and the R^2 is only about .07 in the several regressions. We think that the probability is high that this security had a spuriously significant OLS linear trend.[30]
3. There are five securities which seem to have had significant nonlinear changes in risk that brought the beta coefficient back close to its original level. American–South African, for example, had a trivial linear coefficient but highly significant quadratic, cubic, and quartic terms which imply large deviations of the beta coefficient during the sample from its beginning and ending values. (This security is a curiousity, for its estimated mean beta coefficient was negative and significant in all regressions). For the five securities in this group, the *minimum* difference between the estimated 4- and 10-variable absolute changes was about 175%. This means, of course, that the 4-variable model completely missed what must have been the complex path taken by the coefficient.
4. The estimated changes in beta coefficients are not always due to capital structure changes. As rough support for this assertion, Table 7 includes estimated changes in the debt-equity ratio between July

> 1962 and July 1965, the beginning and ending sample points. These are market values whenever possible. Normally, the debt/equity ratio is the market value of the bonds plus the book value of nonmarketable debt divided by the market value of equity for a given date.

For reasons which we cannot explain, there were no t ratios in excess of 3.5 for the intercept polynomial coefficients of any security. This is mild evidence against both the Sharpe-Lintner and Black versions of the capital asset pricing model because a significant movement in the beta coefficient should be reflected in an opposite and significant movement in the intercept unless changes in the riskless return or in the zero-beta return just happened to be offsetting. The riskless rate or the zero-beta return cannot offset the risk coefficient movements of *all* securities because (1) some securities had implied negative and some had positive changes; and (2) some of the negative changes were associated with high beta coefficients and vice versa.[31] Coupled with this fact is a danger that sampling variation was responsible for the selection of some of these securities with highly varying parameters. Thus, we do not want to assert that a definitive test of any theory has been provided by these last results. In the next section, portfolios will be used to provide a more powerful examination of models which require the intercept to move in a direction opposite to the slope.

Another seeming anomaly is that none of the 84 securities listed on the American Stock Exchange had t ratios as high as 3.5 for any polynomial coefficients. However, if the probability of a security having such a coefficient is equal to the observed frequency, 18/930 = .00194, the probability is roughly 10% that none would be observed in a random sample of 84. This probability would not justify an inference that NYSE- and AMEX-listed securities were truly different.

G. Discriminating Between Aggregate and Individual Nonstationarity

We will now discuss additional tests for the existence of nonstationarity in the overall universe of available data. In Sec. IV.D we demonstrated that our expositional vehicle, a sample of returns on individual assets, did indeed display some kind of nonstationarity. In the application of our method to an individual series, there would be no need to go beyond Section IV.D. Equation (11) would yield as much information about the nature of the nonstationarity as the single series could disclose. In other applications, however, as with our data here, the availability of multiple time series might permit a finer degree of inference about any possible nonstationarity common to all the series. This could be of considerable practical interest. For example, it might not be very interesting to a portfolio manager that some stocks were departing in one direction from previously estimated parameters while other stocks were departing in the opposite direction; but it would be highly interesting to know if there were a common change in the parameters which might signify changes in general market conditions, (e.g., increased general uncertainty).

With many time series, inference about aggregate stationarity and discrimination between individual and aggregate nonstationarity is indeed possible. Going back to the data we have been using here, their aggregation would imply the formation of portfolios of assets and the returns on these portfolios would become the object of our aggregate empirical enquiry.

There are actually two advantages to using portfolios rather than individual securities in data analysis of returns. The first advantage is well known in the asset literature, going back to the original work of Blume (1970) and being refined by Black et al. (1972) among others. This is to average out, by a cross-sectional collection, events peculiar to individual securities. The structure of any data-generating process can be discerned more clearly after such an aggregation because the signal-to-noise ratio is higher. Of course, some care must be taken to assure that the portfolios formed actually differ a priori in their characteristics. This is the same thing as saying that the experiment should be designed to maximize the cross-sectional variation in the parameter of interest.

The second advantage of portfolio formation, which is noted here for the first time to our knowledge, is to smooth out *changes* in the model's parameters which are unique to individual firms. We might find, for example, that the dispersion of the disturbance term for a portfolio is much more stationary than for individual securities. On the other hand, changes in the market price of risk or in the riskless rate of interest would be common to all firms and would thus affect portfolios of any size to the same degree. Since individual parameter nonstationarity would tend to wash out, the ability to perceive common nonstationarity must be enhanced by the portfolio aggregation.

The number of portfolios formed is a matter of judgment. Since the total number of securities is fixed (in our case at 930), the larger the number of securities included in each portfolio, the smaller must be the total number of portfolios. Ideally, one would like to have a very large number of portfolios of very large sizes. We decided to partition the sample into 15 equal-sized portfolios of 62 securities each. Securities were assigned to portfolios based on the relative rank of the average beta coefficient, $\hat{b}_{0j}$ in model (11), from the 4-variable robust regression,[32] estimated over the 160 weeks from July 5, 1962 through July 22, 1965.

If $\hat{b}_{0j}$ was the jth largest of these 930 coefficient estimates, the security was assigned to the INT (j/63) + 1st portfolio, [where INT (x) is the largest integer less than or equal to x]. Thus, portfolio 1 was composed of the 62 securities with the smallest estimates of the beta coefficient, portfolio 2 of the 62 next smallest, etc. Some characteristics of the securities within each portfolio are given in Table 8.[33]

The dispersion of coefficients within each portfolio is quite uniform across portfolios except for the two extremes. This is to be expected, of course, since the extreme portfolios catch the tails of the cross-sectional distribution. The median coefficient in each portfolio differs from the mean only in the third significant digit except for portfolio 1, whose median $\hat{b}_0$ is .285. For portfolio 15, the median is 1.88, and for most other portfolios the median agrees with the mean to the fourth digit.

The portfolios thus formed were followed through a second sample of 160 weeks, beginning August 5, 1965 and ending August 22, 1968.[34] This permits a direct replication of tests for the *existence* of nonstationarity.

For these portfolios and two sample periods, Table 9 reports some of the results of the regression calculations. A full set of results for each period, analogous to Table 2 for individual securities, is available on request. For reasons of space, it did not seem worth reproducing here.

In addition to a small drop in the average explanatory power of all regressions from the first to the second period, there was a wholly anticipated decline in the cross-sectional dispersion of estimated beta coefficients. In the 4-variable robust model, for example, the minimum and maximum estimates of $\hat{b}_0$ were .274 and 1.92 respectively for 1962–1965[35] and .589 and 1.52 respectively for 1965–1968. This is of course attributable to the fact that some securities were incorrectly ranked, and thus assigned to inappropriate portfolios, because of sampling errors, during the first subperiod. Low-ranking portfolios tended to contain a disproportionate number of securities with negative sampling errors and vice versa for high-ranking portfolios. Indeed, this is the very rationale for the technique: to obtain an *unbiased* estimate of risk for each portfolio during the second period while maintaining a reasonable degree of cross-sectional dispersion.

The only noteworthy and unexpected result in Table 9 is the large decline in significance for beta coefficients estimated with the 10-variable model during the second period. Since this is not matched by a corresponding reduction in R^2, it must imply that nonlinear nonstationarities have been responsible for a significant proportion of the observed variation in portfolio returns.

On the subject of nonstationarity in the dispersion of the disturbance term of the simple model, the portfolio results offer some interesting evidence. For individual securities, the robust method determined that

Table 8. Estimated Average Beta Coefficients[a] by Assigned Portfolio, 1962–1965

Portfolio	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
Mean $\hat{b}$	.206	.438	.535	.630	.719	.801	.885	.968	1.06	1.14	1.22	1.33	1.46	1.62	1.91
Cross-sectional Standard Deviation of b_0	.219	.0300	.0256	.0245	.0236	.0264	.0220	.0241	.0253	.0223	.0271	.0320	.0383	.0578	.157

[a] For the 4-variable robust model.

Table 9. Regressions on Portfolios for Two Sample Periods of 160 Weeks Each

No. of Variables	*OLS* 1962–1965	*Robust* 1962–1965	*OLS* 1965–1965	*Robust* 1965–1965
	Median[a] value of t ratio for average risk coefficient $\hat{b}_0$			
2	22.3	22.4	18.3	19.7
4	21.2	21.5	18.2	19.7
10	18.6	19.2	12.4	12.8
	Median[a] coefficient of determination, R^2			
2	.756	.759	.678	.714
4	.756	.760	.675	.712
10	.757	.765	.679	.713

[a] The average of the seventh and eighth largest sample values out of 15.

about 3.0% of the observations were outliers. (The average number of nonexcluded observations was 155 out of 160 possible.) For the portfolio calculations, only 30 outliers in total were computed in each subperiod (out of 2400 possible). This is only 1.25%, less than half the percentage of outliers found with individual securities.[36] An implication follows that a substantial part of the observed thick-tailed non-Gaussian nature of the distribution of residuals from the individual regressions is due to nonstationarity and not to a basic underlying non-Gaussian stable law. Since there is very little difference between robust and ordinary least squares calculations for portfolios, the more straightforward sampling properties of the OLS methods can be relied upon.

Table 10, a companion to Table 3, reports a test for the *existence* of nonstationarity. It contains the cross-sectional mean t values of the higher-order polynomial coefficients of model (11) divided by their estimated standard errors. Only the robust estimates are used because the sampling distribution is known to follow the Student law in this case, (see Section IV.D) under the null hypothesis of stationarity. Coefficients which are significantly different from zero provide evidence that the basic model did not have stationary parameters during the sample period.[37] The table also reports the mean coefficients divided by their standard errors as computed from the cross-sectional standard deviations of the coefficients ($\hat{\sigma}_c$).

Table 10 indicates that nonlinear terms in the 10-variable model displayed the most measured significance during the second period. Although the average estimated linear trend in the 4-variable model was statistically significant, it was not large (.041) and was neither as large

Table 10. Tests for the Presence of Nonstationary Parameters in 15 Portfolios Formed from 930 NYSE and AMEX Securities, 1962–1968 (Robust Estimates)

Estimated Parameter	No. of Variables in Model	1962–1965 $\bar{t}/(\hat{\sigma}_t/\sqrt{N})$ (t Ratio)	1962–1965 $\bar{c}/(\hat{\sigma}_c/\sqrt{N})$ (Coefficient)	1965–1968 $\bar{t}/(\hat{\sigma}_t/\sqrt{N})$ (t Ratio)	1965–1968 $\bar{c}/(\hat{\sigma}_c/\sqrt{N})$ (Coefficient)
Intercept Trend, $\hat{a}_1$	4	3.72*	6.91*	.0175	− 1.13
Slope Trend, $\hat{b}_1$	4	−1.56	− 1.85	2.44*	3.88*
Intercept Trend, $\hat{a}_1$	10	4.88*	8.34*	− 1.99*	− 3.88*
Intercept Quadratic, $\hat{a}_2$	10	−1.35	− 1.04	2.94*	5.28*
Intercept Cubic, $\hat{a}_3$	10	−6.77*	−15.3*	− 8.20*	−15.2*
Intercept Quartic, $\hat{a}_4$	10	−5.96*	−11.5*	3.98*	8.94*
Slope Trend, $\hat{b}_1$	10	− .901	− 1.00	5.17*	11.0*
Slope Quadratic, $\hat{b}_2$	10	5.69*	10.1*	1.16	1.35
Slope Cubic, $\hat{b}_3$	10	.824	2.43*	10.5*	15.0*
Slope Quartic, $\hat{b}_4$	10	−2.96*	− 3.78*	− 7.06*	−13.4*

* Rejection of the hypothesis of stationarity at the 95% level of significance.

in absolute value nor as significant as the nonlinear coefficients in the 10-variable model.[38]

There appears to have been significant changes in parameters during both subperiods but there seems to be little connection between the two. Of the 10 pairs of coefficients computed, only 2 pairs are significant with the same sign and the remaining pairs have either one nonsignificant member or are significant with opposite signs.

In comparing the portfolio t ratio results for the first subperiod (1962–1965) with the same computations for individual securities (the next-to-last column of Table 3), the portfolio coefficients are smaller in seven of the eight cases which were "significant" for individual securities. However, in only two cases did the estimated coefficient become "insignificant" at the 95% level and both of these cases retained the same sign. As we argued in Sec. IV.D, this test statistic will tend to overstate the significance of nonstationarity if there is positive dependence cross-sectionally in the estimated coefficients. The comparison of

Tables 3 and 10 strongly suggests that such positive dependence is indeed present.

The correction of this bias turns out to be a rather difficult task because there is no a priori information about how the covariance itself might differ across pairs of securities (or portfolios). In order to explain our solution to this problem as simply as possible, let us use only the simple two-variable ordinary least-squares model in deviation form. (The discussion generalizes readily but tediously to the more complex models and the calculations were actually made for them).

With the simple model in deviation form, we have for portfolio j

$$R_{j,t} - \bar{R}_j = b_{0,j}(R_{m,t} - \bar{R}_m) + \varepsilon_{j,t}$$

where the bars indicate intertemporal means [cf. Eq. (1)]. If, for compactness, we denote $x_t = R_{mt} - \bar{R}_m$ and the sample vectors $x' \equiv [x_1; x_2; \ldots; x_T]$ and $\varepsilon_j' = [\varepsilon_{j,1};\varepsilon_{j,2}; \ldots; \varepsilon_{j,T}]$ where T is the sample size, then the least-squares estimate of $b_{0,j}$ is given by

$$\hat{b}_{0,j} = b_{0,j} + (x'x)^{-1}x'\varepsilon_j.$$

Now assume that each *true* coefficient (not its estimate) is generated by deviations about a mean coefficient, that is,

$$b_{0,j} = b_0 + \xi_j$$

where b_0 is the mean cross-sectional coefficient, whose estimation is the object of our procedure, and ξ_j is a stochastic error for which we assume standard spherical properties. Thus, the estimated coefficient can be considered the dependent variable in a regression function with no explanatory variables:

$$\hat{b}_{0j} = b_0 + \mu_j \qquad j = 1, \ldots, 15$$

where

$$\mu_j \equiv \xi_j + (x'x)^{-1}x'\varepsilon_j.$$

The problem in estimating b_0 arises because the disturbances of this model, the μ terms, are not necessarily independent. However, one can observe from the model's structure that an estimate of the complete covariance matrix of its disturbance, $E(\mu\mu')$, can be obtained from a combination of the cross-sectional variance in the estimated coefficients plus the estimated covariance matrix of residuals from the original regressions. In other words, making the reasonable presumption that ξ_j and ε_j are uncorrelated,

$$E(\mu_j\mu_k) = E(\xi_j\xi_k) + (x'x)^{-1}x'E(\varepsilon_j\varepsilon_k')x(x'x)^{-1}$$

with the additional presumption that lagged cross-correlations of the residuals are uncorrelated [that $\mathrm{Cov}(\varepsilon_{j,t}, \varepsilon_{k,t-i}) = 0$ for $i \neq 0$], this expression reduces to

$$E(\mu_j \mu_k) = E(\xi_j \xi_k) + (x'x)^{-1} E(\varepsilon_j \varepsilon_k')$$

or in vector notation

$$E(\mu\mu') = E(\xi\xi') + (x'x)^{-1} E(\varepsilon'\varepsilon)$$

where ε is the (160×15) matrix of disturbances from the 15 separate regressions for individual portfolios [and thus $E(\varepsilon'\varepsilon)$ is the concurrent cross-covariance matrix of disturbances].

Since ξ is by assumption a standard spherical error, the off-diagonal elements of $E(\xi\xi')$ are zero and all diagonal elements are the same constant, denoted σ_ξ^2. An estimate of σ_ξ^2 is given, therefore, by

$$(\hat{\sigma}_\xi^2) = (1/15)^2\, 1'\hat{u}'\hat{u}1 - 1'(x'x)^{-1}\hat{\varepsilon}'\hat{\varepsilon}1$$

where 1 is the 15-element unit vector, $\hat{\varepsilon}$ is the matrix of *residuals* from the 15 separate portfolio regressions, and $\hat{u} \equiv \hat{b}_0 - \bar{\hat{b}}_0$ is the vector of deviations of coefficients about their cross-sectional mean, $\bar{\hat{b}}_0$.

Given this estimate of the cross-sectional variance of the true coefficient, an estimate of the covariance matrix of disturbances in the regression model above is

$$\hat{\Omega} = \hat{E}(\mu\mu') = \hat{\sigma}_\xi^2 I + (x'x)^{-1}\hat{\varepsilon}'\hat{\varepsilon}$$

where I is the 15×15 identity matrix. Then an estimate of the true mean beta is obtained from the Aitken estimator (or by generalized least squares) as

$$\bar{b} \equiv (1'\hat{\Omega}^{-1}1)^{-1}\,(1'\hat{\Omega}^{-1}\hat{b}_0)$$

and the standard error of $\bar{b}$ is obtained as

$$\hat{s} = \sqrt{(1'\hat{\Omega}^{-1}1)^{-1}}.$$

Generally speaking, this estimate of b_0 can and does differ from the simple cross-sectional mean of the $\hat{b}_0$ terms (which we denoted $\bar{\hat{b}}_0$ and used to estimate the disturbances $\hat{u}$ above). It is therefore natural to reestimate the $\hat{u}$ terms using this new and presumably more precise estimate of b_0. We did this and then iterated the procedure until the estimate $\bar{b}$ changed by less than .01%. This was accomplished in at most eight iterations. The resulting estimates and their t ratios using the standard error calculated as above (but with the final iteration) are given in Table 11. These t ratios are in principle accurate measures of significance for the average coefficient after taking into account the cross-sectional covariation.

Table 11. Tests for the Presence of Nonstationarity, Aitken-estimator Correction for Cross-sectional Covariation

Estimated Parameter	*No. of Variables in Model*	*1962–1965*	*1965–1968*
		t Ratios from Aitken Procedure	
Intercept Trend	4	.528	.481
Slope Trend	4	−2.04*	1.25
Intercept Trend	10	1.47	− .853
Intercept Quadratic	10	−1.22	.925
Intercept Cubic	10	−3.12*	−2.37*
Intercept Quartic	10	−2.87*	.993
Slope Trend	10	−1.75*	1.92*
Slope Quadratic	10	4.11*	1.02
Slope Cubic	10	− .267	2.92*
Slope Quartic	10	−2.16*	−2.17*
	Coefficients and t Ratios from Aitken Procedure		
Mean Slope, $\hat{b}_0$	4	1.02	.999
		(8.92)	(12.5)
Mean Slope, $\hat{b}_0$	10	1.00	.945
		(8.93)	(13.3)

* Rejection of the hypothesis of stationarity of the 95% level of significance.

The general impression of these results, on comparing them with Table 10, is a reduction in the significance level. With one or two exceptions, the signs remain unchanged but the absolute size of the estimated t ratio for the higher-order polynomial coefficients is reduced. Nevertheless, there are still indications of significant nonstationarity in both periods. The slope trend estimates for the 1962–1965 period have actually increased in significance, marginally, in both the 4-variable and 10-variable models. We can be fairly certain that these results are purged of all bias introduced by cross-sectional dependence. Thus an inference of significantly concurrent nonstationarity in the coefficients of model (1) would be very reasonable.

H. Using Aggregated Data to Detect Presence of an Underlying Common Generating Process

As a final topic in this chapter, we wish to present a suggestion for using the presence of nonstationarity to examine the structure of the data-generating process. Imagine that all of the time series available to an investigator have been generated by an underlying process of identical form but with different parameters. Imagine further that the parameters wander over time but that their paths are tracked by the methodology outlined in the preceding sections. We assert that these estimated time paths give an opportunity to infer something about the generating process, if it is indeed common. Intuitively, the cross-sectional relationship among the tracked parameters at a *given* time should have as its coefficients the values of the factors in the underlying process at that same time. If the form of this cross-sectional relation were not stationary, the investigator would be obliged to deny the validity of the particular generating process he had hypothesized.

As a concrete example, we now use the cross-sectional sample of 15 portfolios in a simple investigation of the two-factor, "zero-beta" generating process, the process in which the coefficient α_{jt} in (1) is equal to $R_{z,t}(1 - \beta_{j,t})$ where $R_{z,t}$ is the return on some zero-beta portfolio. This model has a very special structure, as it implies that the intercept estimate from (1) will be given by

$$\hat{\alpha}_{j,t} = d_0 + d_{1,t}(1 - \hat{\beta}_{j,t}) \tag{12}$$

where $\hat{\alpha}_{jt}$ and $\hat{\beta}_{jt}$ are estimates of (8) and (9) for portfolio j in time t. An appropriate procedure would be to fit (12) for *any* value of t and test the two-factor null hypothesis: $E(\hat{d}_0) = 0$, $E(\hat{d}_{1,t}) = R_{z,t}$. The second part of the hypothesis is not easy to test because $R_{z,t}$ is not directly observable. Black et al. (1972) give an estimating method but their estimate has the property that $E(\hat{d}_0) = 0$, even when the true value of d_0 is nonzero. One way to avoid this problem is to form a long-term average of $\hat{\alpha}_{jt}$ and $\hat{\beta}_{jt}$ and then estimate (12) using these averages without specifying $R_{z,t}$ at all (letting *its* average value be estimated by $\hat{d}_1$).

If the averaging period is chosen equal to the range of orthogonality for the polynomial estimates, then the average value of $\hat{\alpha}_{jt}$ will be $\hat{\alpha}_{0j}$ from (11) because the average polynomials of order greater than 1 will be zero. The second term on the right side of (12) will consist of two parts: $\bar{d}_1$ (which will be equal to $\bar{R}_z$ if the hypothesis is true) and $-\overline{d_1\hat{\beta}_j}$, which is the intertemporal average value of $-\beta_{j,t} \cdot R_{z,t}$ under the null hypothesis. If the market is efficient, the zero-beta portfolio return should fluctuate unpredictably and thus not be correlated over time with the fluctuations in $\beta_{j,t}$ (the risk of a particular nonzero-beta portfolio). If this condition is satisfied, $E(\beta_{j,t} \cdot R_{z,t}) = E(\beta_{j,t})E(R_{z,t})$. But when the period is chosen such that the higher-order polynomials are orthogonal, $E(\beta_{j,t})$ is just $\hat{b}_{0j}$, the zero-order polynomial coefficient. This finally provides a testable cross-sectional model

$$\hat{a}_{0j} = \hat{\delta}_0 + \hat{\delta}_1(1 - \hat{b}_{0j}) + \varepsilon_j \tag{13}$$

where $E(\hat{\delta}_0) = 0$ [and $E(\hat{\delta}_1) = \bar{R}_z$] if the two-factor generating process is valid.

The estimated coefficients of (13) will be asymptotically biased because the explanatory variable contains an error, being an estimate itself. There are two possible ways to alleviate this: (1) we could use the estimated regression standard error of $\hat{b}_{0j}$ to correct the bias; or (2) we could use unstrumental variables to obtain consistent estimates. Table 9 shows that the t ratio for $\hat{b}_{0j}$ is on the order of 20.[39] Since the mean value of $\hat{b}_{0j}$ is on the order of unity, the standard error is on the order of .05. (Over the 6 models and 15 portfolios, the actual maximum standard error was .0818.) Since the cross-sectional standard deviation of $\hat{b}_{0j}$ is on the order of .25, the *maximum* percentage decline in the absolute value of $\hat{\delta}_1$ caused by the error in $\hat{b}_{0j}$ is around 10.1% and it will probably be less than half this amount. The bias in the intercept of (13) is likely to be much smaller than the bias in the slope. In fact, the bias in $\hat{\delta}_0$ would be exactly zero if the cross-sectional average of $\hat{b}_{0j}$ were exactly unity.

Results for model (13) applied to the cross-section of 15 portfolios are given in Table 12. Two estimates are given for each coefficient, the first being computed by ordinary least squares and the second using instrumental variables. As instrumental variables, we used estimates of $\hat{b}_{0j}$ for the period 1962–1965. These were natural instruments since they were computed for the same portfolios but in an earlier period.

The results indicate a highly significant nonzero value for $\hat{\delta}_0$ which is strong evidence against the two-factor zero-beta market model. The estimates of δ_0 obtained with the instrumental variables are very close to those obtained with the classical method.

As for the slope, supposedly an estimate of the mean zero-beta portfolio return, it is significantly negative in all cases. Notice, however, that the effect of the error in the explanatory variable shows up in the difference between ordinary least-squares and instrumental variable estimates. In every case, the instrumental variable estimate is greater than or equal to the OLS estimate in absolute magnitude. The maximum difference is 14.8% for the 10-variable model using the OLS method to estimate $\hat{b}_{0j}$.[40] This conforms very well to the anticipated errors-in-variables bias.

The highly significant intercept cannot be explained by errors in variables. In the 10-variable models, the bias would actually work in the wrong direction since in these cases the average value of $1 - \hat{b}_{0j}$ is positive (and thus an error adjustment in the slope $\hat{\delta}_1$, making it smaller, would move $\hat{\delta}_0$ even further from zero). For the 2- and 4-variable models, even if the error in $\hat{\delta}_1$ were 25% (and it is certainly much less than this), the decline in the estimated intercept would be only on the order of .1, leaving it essentially unchanged and still over 20 standard errors from zero.

We can only conclude that the stochastic process generating asset

V. CONCLUSIONS

We have presented and explained a methodology for using linear models that display nonstationary parameters. The time paths of slope and constant terms were estimated by orthogonal polynomials of time. Nonstationarities in the distribution of the disturbances were expurgated by the technique of robust regression. The method was applied to the one-factor market model (MM1) of asset returns (1) and the principal empirical results were as follows:

1. There is substantial evidence that MM1's disturbances are non-Gaussian and there is good evidence that this is caused, at least partly, by nonstationarity over time in the distribution of disturbances. The methodology handled this problem adequately in the sense that it provided parameter estimates which followed closely the Gaussian law.
2. There is strong evidence that MM1 has nonstationary parameters. The principal type of nonstationarity (or better said, the most significant type) seems to be common across all securities. This is consistent with the hypothesis that there are significant factors omitted by the simple model. There is some evidence also that nonstationarity is present on the individual security level. Since this could not be caused by the omission of a common factor, it must be due to changes in the characteristics of individual firms.
3. There have been episodes of unusual activity on the New York and American Stock Exchanges. During our sample period, 1962–1965, the number of outliers (unusual observations) clustered in the early weeks of the record; and there was strong intertemporal dependence in the number of outliers. This did not seem to be caused by variations over time in the market index but rather by comovements across securities in the distribution of disturbances.
4. The method was able to identify securities with particularly large changes in parameters over time. This was true of all parameters. Securities that have displayed extremely large changes in the portfolio risk coefficient over time were identified by name, as were securities which had unusually large numbers of outliers (and thus displayed strong nonstationarity in the disturbances or marked departure from Gaussian disturbances).
5. Refined investigations of nonstationarity and investigations of the data-generating process were presented with aggregated data as examples of extensions of the basic method to multiple time series. Our method can be employed profitably under many circumstances as a data-screening device, for the identification of individual cases with particular nonstationarities, or to test hypotheses which purport to predict such changes.

Table 12. Tests of the Zero-Beta Two-factor Process with 15 Portfolios, 1965–1968

Number of Variables:	2	4	10	2	4	10
Method:	*OLS*			*Robust*		
δ_0, Intercept (%/year)	17.5	17.5	21.1	17.7	17.7	21.3
(t)	(27.5)	(27.7)	(26.1)	(21.0)	(21.0)	(22.9)
δ_0, Intercept[a]	17.5	17.5	21.1	17.7	17.7	21.4
δ_1, Slope (%/year)	−6.64	−7.15	−12.0	−10.1	−10.6	−15.5
(t)	(2.70)	(−2.95)	(−3.61)	(−3.05)	(−3.23)	(−4.01)
δ_1, Slope[a]	−6.83	−7.29	−14.1	−10.1	−10.7	−17.4
Adjusted R^2	.311	.355	.462	.372	.403	.519
Studentized Range of Residuals	3.55	3.56	3.77	3.77	3.80	3.88

[a] Using as instrumental variables the corresponding estimated coefficients for 1962–1965 (see text).

returns during 1965–1968 on the New York and American Stock Exchanges was definitely not the two-factor, zero-beta market model with the index we used. (Recall that it was the Standard & Poor's 500 Index.) The principal reason for these results is the exceedingly bad performance of this index. While randomly selected portfolios had average monthly returns in the neighborhood of 25% per annum during this sample period, the S&P 500 had an average monthly return of only about 9.0%. Since the average risk coefficient of a randomly chosen portfolio was near unity (the average $\hat{b}_{0j}$ for the 15 portfolios used in this paper ranged between .995 and 1.02 over the six regressions computed), the poor performance of the S&P Index cannot be attributed to a lower level of "risk" as measured by b_0. This "risk" measure was not lower. Thus, our test of the two-factor generating process is equivalent to the finding that a particular well-diversified portfolio (the S&P Index) did not provide returns commensurate with its level of risk.[41]

It is possible to use the entire time paths of estimated coefficients including the higher-order estimates of changes in parameters in similar computations. In fact, if it can be presumed that $R_{z,t}$, the zero-beta value, follows a random walk with constant drift, then the two-factor generating process implies that *any* of the higher-order coefficients in the equation,

$$a_{ij} = \delta_{i0} + \delta_{i1} b_{ij} + \varepsilon_{ij} \qquad i \geq 1 \tag{14}$$

should satisfy the hypothesis $E(\hat{\delta}_{i0}) = 0$, and $E(\hat{\delta}_{i1}) = -\bar{R}_z$. (See note 42 for proof.) Due to space considerations, we do not report these empirical results.

ACKNOWLEDGMENTS

We would like to thank the members of the University of Wisconsin and University of Pennsylvania Departments of Statistics, the Center for Operations Research and Econometrics at the University of Louvain, the Graduate Schools of Business at the University of British Columbia and at the University of California at Los Angeles, and the members of the finance workshop at the European Institute for Advanced Studies in Management. We also benefited from the detailed comments of Anton Barten and Eugene Fama. Edward Beauvais provided many useful suggestions as well as research assistance. Professor Hinichs research was supported by the Office of Naval Research (Probability and Statistics Program).

NOTES

1. See, for example, Francis and Archer (1971, chap. 5), Sharpe (1970, 1980), Mossin (1973), Fama and Miller (1972), Black (1972).

2. Some of the events which have been studied are stock splits, secondary offerings, dividend changes, accounting manipulations, and mergers. Brealey (1971) gives a reference list and other examples.

3. See, for example, the theoretical papers by Chen (1979), Huberman (1979), Jarrow and Rudd (1980), Garman and Ohlson (1978, 1979), and the empirical paper by Gehr (1978) and Roll and Ross (1980).

4. For an extensive analysis of this factor's influence, including empirical testing, see Beaver et al. (1970), Hamada (1972), Boness et al. (1974), and Blume (1975).

5. See Arrow (1971, chap. 3).

6. There may be other influences on β_j. Rosenberg and McKibben (1973) list over 30 and present empirical evidence that at least some of them actually have been influential. Schaefer et al. (1974) present alternative models of the β_j as a stochastic process. Blume (1975) found a tendency for extreme values of β to disappear over time.

7. Rosenberg and McKibben (1973) found six accounting measures, and two historical measures, which seem to have had significant influence on the variance of ε_j.

8. Evidence on the non-Gaussian nature of asset returns seems to have been given an explicit (and thorough) treatment first by Mandelbrot (1963). Other evidence has accumulated over time. See, e.g., Fama (1965), Roll (1970), Schwartz and Altman (1973).

9. Press (1967), Mandelbrot and Taylor (1967), Officer (1972), Clark (1973), Barnea and Downes (1973), Hsu et al. (1973), Blattberg and Gonedes (1974) offer alternative explanations to the basic phenomena that asset returns are too thick-tailed to have been generated by a stationary Gaussian process, i.e., that they have too many extreme observations. Data errors are another source of econometric troubles which may result in seemingly thick-tailed residuals from an ordinary least-squares fit of model (1). In a recent paper, Rosenberg and Houglet (1974) present striking evidence on the effect of errors and suggest the use of "a statistical method that minimizes the effect of outliers," by discarding or truncating them or by using a procedure "more 'robust' than the quadratic methods" (pp. 1308–1309). As we shall see, the methodology employed in this chapter has exactly this aim.

10. The characteristic function for a symmetric stable distribution with zero location, scale σ, and characteristic exponent γ is:

$$\phi_\varepsilon(t) = \exp(-|\sigma t|^\gamma).$$

11. For details see Hinich and Talwar (1975).

12. Fama et al. (1969).

13. Kaplan and Roll (1972).

14. Scholes (1972).

15. Watts (1973). In this case, the size of the nonzero mean disturbance seems to be quite small.

16. A number of other events, as well as some of those already mentioned above, are summarized in Brealey (1971).

17. For example, if the observed period included two years of weekly data (104 weeks), the time variable would run −1, −51/52, −50/52, ..., 51/52, 1.

18. See Abramowitz and Stegun (1964, chap. 22).

19. We wish to thank the Management Science Department of Wells-Fargo Bank for making the data available in a covenient form.

20. The Thursday closing prices of each week were used. If Thursday was a holiday, no observation was used for that week.

21. We began the study in Pittsburgh with an earlier version of the tape. At that point, we had 1042 stocks in a sample selected similarly. On continuing our work in Belgium on the same tape, the records for only about 100 securities could be read. Evidently, the tape had been damaged in transport. Wells-Fargo kindly supplied another copy. On this tape, we obtained the sample size of 930.

It is interesting to note that a number of securities which had been read successfully on the Pittsburgh system and which the Belgian system designated as faulty records on the second tape were securities with a very large number of outliers computed by the robust technique. As an example, Benguet had 87 outliers in the first 160 observations in the Pittsburgh runs. The Belgian system rejected Benguet as a bad record. We think this is evidence that the original tape, from which both tapes were copied, actually had a bad spot in the Benguet record that is acceptable on some systems. It is quite likely that Benguet's data record is free of error in its original form, but when tapes are copied and recopied, shipped, and used frequently, physical damage can occur. This suggests that the robust technique is particularly valuable for large data banks that are distributed widely. Even if a given machine installation accepts the physically damaged record as valid, the technique may reject it as an outlier.

One other disquieting piece of evidence: In the ordinary least-squares runs in Pittsburgh, the average standard deviation of regression residuals across 1042 securities was 10 times as high as the average for 930 securities in the Belgian runs. We later discovered that this was due to a single security! The mean standard deviations for the robust cases differed only in the third decimal place in the two runs. These comments are very similar in spirit to Rosenberg and Houglet's (1974).

22. In fact, most choices in empirical work are made in such a sequential and stochastic manner; even when the final paper describes the choices made as having been preordained by logic. The actual procedure for sample selection and consolidation really makes no difference provided that the results observed at an intermediate step do not influence the results reported in the final paper.

23. The formula for the asymptotic standard deviation of a sample range is obtained from known formulas for asymptotic variances and covariances of order statistics. See, e.g., Cramer (1946, p. 369).

24. The logarithm was taken in order to reduce skewness in the dependent variable.

25. The *mean* t values for the coefficients in Table 6 are as follows:

Technique	*No. of Variables*		
	2	*4*	*10*
OLS	3.23	2.94	2.42
ROB	3.39	3.24	2.72

26. That is, estimates with higher t values.

27. It is *twice* the slope because time has been normalized to span -1 to $+1$.

28. In all cases, the 4-variable and 10-variable estimates agreed on the level of significance.

29. The t ratios are on the order of -1.4.

30. One other piece of supporting evidence for this assertion is the sizable difference, 1.245 versus .686, between the beta coefficient estimated (with the 2-variable model) by OLS and robust, respectively. (The t ratios were 5.12 and 3.52, respectively.) This strongly suggests a bad and very influential sample observation.

31. All four cases are present among the 18 securities of Table 7. For examples:

Security	*Direction of Risk Coeff. Change*	*Estimated Mean Risk Coeff. (10-Vbl. ROB)*
Amer. Comm. Lines	+	.801
Zapata-Norness	+	1.74
Florida Power	–	.525
W. R. Grace	–	1.25

32. Using b_{0j} from the 4-variable robust model is also a somewhat arbitrary choice. We believe, however, that the robust estimators are better than the OLS estimators and that some nonstationarity is present in the data. Thus, only the 4- and 10-variable robust estimates of b_{0j} would be good candidates for the portfolio assignment criterion. The final choice of the 4-variable rather than the 10-variable estimate was completely arbitrary but the similarity in these estimates implies that the choice did not matter.

33. The portfolios were formed with an equal weighting on each security's return in every period. This is known as an equally weighted, rebalanced portfolio formation method.

34. During the second subsample, some securities had missing observations. When this occurred, the portfolio for that week was composed of an equally weighted average of the securities which *did* have observed returns. In no case was a portfolio composed of fewer than 57 securities (out of 62 possible).

35. N.B. The estimated risk coefficients for a portfolio can differ from the mean estimates for its component securities with the robust model. This is because the outlier observations can be and were different when performing the robust calculation on individual securities and on aggregates. Thus, for example, the cross-sectional mean *individual* estimates for 1962–1965 were minimum, .206, and maximum, 1.91 (see Table 8).

36. The difference between the two results is highly significant. For the individual securities, given that the number of outliers is computed from a true random sample, the standard error of the sample proportion is on the order of .0004. For portfolios, the associated standard error is about .002. Thus, the difference in proportions is at least 8 times its standard error.

37. Another test included in Table 3 involved the cross-sectional 90-percentile range. (This was a test for nonconcurrent nonstationarity among individual time series.) It was not repeated here because there are only 15 values (portfolios) in the cross-sectional distribution and order statistics are only asymptotically Gaussian. Besides, it is quite likely that changes in parameters which are nonconcurrent across securities would cancel out in the process of portfolio aggregation. Thus, in all probability, they would have been undetectable anyway in the portfolio data, even if they were present in individual security data.

38. Average cross-sectional robust estimates of b_3 and b_4 for the 15 portfolios during the second period were .340 and $-.320$, respectively. These are the coefficients of the cubic and quartic polynomials of the change in the estimated risk coefficient, denoted "slope cubic" and "slope quartic" in Table 10.

39. For the second subsample, however, it is only on the order of 12 for the 10-variable model.

40. The differences between ordinary estimates of δ_1 and instrumental variable estimates suggest that the 10-variable model does not provide as accurate estimates of the $\hat{b}_{0i}$ as the 2- and 4-variable models. However, the closer agreement of the δ_1 estimates using the robust method relative to the OLS method again shows the robust method's superiority.

41. This is not all that surprising since indexes which may seem to have very similar properties can give quite different results for model (1). See Roll (1977) for a more complete analysis of the index's consequences.

Another possibility, however, is bias in estimated risk parameters arising from non-synchronous trading. See Dimson (1979). We have not had a chance to investigate this possible explanation.

42. According to the two-factor generating process, the intercept and slope coefficients of (1) are given by

$$\alpha_{j,t} = (1 - \beta_{j,t})R_{zt}. \qquad (12')$$

Define the vector P_t containing as elements the orthogonal polynomials of Table 1 evaluated at time t. That is,

$$P_t' = [P_0(t):P_1(t): \ldots :P_n(t)]$$

where n is a number sufficiently large to guarantee a good approximation such that

$$\alpha_{j,t} \approx a_j' P_t$$

$$\beta_{j,t} \approx b_j' P_t$$

contain only trivial errors and where a_j and b_j are column vectors of constants whose estimation is the object of the empirical fitting,

$$a_j' \equiv [a_{0j}:a_{1j}: \ldots :a_{nj}], \qquad b_j' = [b_{0j}:b_{1j}: \ldots :b_{nj}].$$

This means that the two-factor generating process (12′) can be written as

$$a_j'P_t = (1 - b_j'P_t)R_{zt} \qquad (13')$$

and we note that since

$$\int_{-1}^{1} P_i(t) = \begin{cases} 2 \text{ for } i = 0 \\ 0 \text{ for } i \geq 1 \end{cases}$$

we can integrate both sides of (13′) to obtain $a_{0j} = (1 - b_{0j})\bar{R}_z$, which is equivalent to the fitted equation (13) in the test (with $\hat{\delta}_0 = 0$ and $\hat{\delta}_1 = \bar{R}_z$). This operation presumes that $R_{z,t}$ is unrelated to $P_i(t)$ $(i \geq 1)$, which is satisfied if $R_{z,t}$ fluctuates randomly.

To obtain a similar test with the full vector of polynomial coefficients, we define the integrated outer product matrix of P_t with itself as

$$\mathbf{P} \equiv \int_{-1}^{1} P_tP_t' \, dt$$

and we note that $\mathbf{P}$ is a positive diagonal matrix. (The integrals of off-diagonal elements are identically zero because the polynomials are mutually orthogonal over the range [−1,1]). Premultiplying (13′) by P_t, rearranging, and integrating over t [while assuming that $R_{z,t}$ and $P_i(t)$ $(i \geq 1)$ are unrelated], results in the equation

$$\mathbf{P}a_j = (\iota' - \mathbf{P}b_j)\bar{R}_z$$

where $\iota' = [2:0:0 \ldots :0]$. Since $\mathbf{P}$ is positive diagonal, its inverse exists and $a_j = (\mathbf{P}^{-1}\iota' - b_j)\bar{R}_z$. But the first element of $\mathbf{P}^{-1}$ is ½, so that

$$a_{0j} = (1 - b_{0j})\,\bar{R}_z$$

and

$$a_{ij} = -b_{ij}\,\bar{R}_z \qquad i \geq 1.$$

This last is Eq. (14) of the text (with $\delta_{i0} = 0$ and $\delta_{i1} = -\bar{R}_z$) while the first equation is again (13) of the text.

REFERENCES

Abramowitz, Milton, and Stegun, Irene A., eds. (1964). *Handbook of Mathematical Functions*. Washington: U.S. Govt. of Printing Office.

Arrow, Kenneth J. (1971). *Essays in the Theory of Risk-Bearing*. Chicago: Markham.

Barnea, Amir, and Downes, David H. (1973). "A Re-examination of the Empirical Distribution of Stock Price Changes," *Journal of the American Statistical Association,* 68 (June), 348–50.

Beaver, William, Kettler, Paul, and Scholes, Myron (1970). "The Association between Market Determined and Accounting Determined Risk Measures," *The Accounting Review,* 45 (October).

Bickel, P. J. (1973). "On Some Analogues to Linear Combinations of Order Statistics in the Linear Model," *Annals of Statistics,* **1** (July), 597–616.

Black, Fischer (1972). "Capital Market Equilibrium with Restricted Borrowing," *Journal of Business,* 45 (July), 444–54.

Black, Fischer, Jensen, Michael C., and Scholes, Myron (1972). "The Capital Asset Pricing Model: Some Empirical Tests," in M.C. Jensen, ed., *Studies in the Theory of Capital Markets*. New York: Praeger, pp. 79–124.

Blattberg, Robert C., and Gonedes, Nicholas J. (1974). "A Comparison of Stable and Student Distributions as Statistical Models for Stock Prices," *Journal of Business,* 47 (April), 244–80.

Blattberg, Robert C., and Sargent, T. (1971) "Regression with Non-Gaussian Stable Disturbances: Some Sampling Results," *Econometrica,* **39** 501–10.

Blume, Marshall E. (1970). "Portfolio Theory: A Step Towards its Practical Application," *Journal of Business,* 43 (April), 152–73.

Blume, Marshall E. (1975). "Betas and Their Regression Tendencies," *Journal of Finance,* 30 (June), 785–795.

Blume, Marshall E., and Friend, Irwin (1973). "A New Look at the Capital Asset Pricing Model," *Journal of Finance,* 28 (March), 19–33.

Boness, A. James, Chen, Andrew H., and Jatusipitak, Som (1974). "Investigations of Nonstationarity in Prices," *Journal of Business,* 47 (October), 518–37.

Brealey, Richard (1971). *Security Prices in a Competitive Market: More about Risk and Return from Common Stocks*. Cambridge, Mass.. MIT Press.

Chen, Nai-fu (1979). "The Arbitrage Pricing Theory: Estimation and Applications." Working Paper, Graduate School of Management, Univ. of California at Los Angeles, October.

Clark, Peter K. (1973). "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica,* 41 (January), 135–55.

Cooley, Thomas F., and Prescott, Edward C. (1973). "An Adaptive Regression Model," *International Economic Review,* 14 (June), 364–71.

Cramer, Harald (1946). *Mathematical Methods of Statistics*. Princeton: Princeton Univ. Press.

Dimson, Elroy (1979). "Risk Measurement When Shares are Subject to Infrequent Trading," Journal of Financial Economics, 7 (June), 197–226.

Fama, Eugene F. (1965). "The Behavior of Stock Market Prices," *Journal of Business,* 38 (January), 35–105.

Fama, Eugene F., and Miller, Merton H. (1972). *The Theory of Finance.* New York: Holt, Rinehart and Winston.

Fama, Eugene F., and Roll, Richard (1968). "Some Properties of Symmetric Stable Distributions," *Journal of the American Statistical Association,* 63 (September), 817–36.

Fama, Eugene F., Fisher, Lawrence, Jensen, Michael C., and Roll, Richard (1969). "The Adjustment of Stock Prices to New Information," *International Economic Review,* 10 (February), 1–21.

Farley, John U., and Hinich, Melvin J. (1970). "A Test for Shifting Slope Coefficient in a Linear Model," *Journal of the American Statistical Association,* 65 (September), 1320–29.

Farley, John U., Hinich, Melvin H., and McGuire, Timothy W. (1975). "Testing for a Shift in the Slopes of a Multivariate Linear Time Series Model," *Journal of Econometrics,* 3, 297–318.

Francis, John Clark, and Archer, Stephen H. (1971). *Portfolio Analysis.* Englewood Cliffs, N.J.: Prentice-Hall.

Garman, Mark B., and Ohlson, James A. (1979). "Information and the Sequential Valuation of Assets in Arbitrage-Free Economies." Working Paper, Department of Business Administration, University of California, Berkeley, January.

Gehr, Adam, Jr. (1978). "Some Tests of the Arbitrage Pricing Theory," *Journal of the Midwest Finance Association,* pp. 91–105.

Hamada, Robert S. (1972). "The Effect of the Firm's Capital Structure on the Systematic Risk of Common Stocks," *Journal of Finance,* 27 (May), 435–52.

Hinich, Melvin J., and Talwar, Prem (1975). "A Simple Method for Robust Regression," *Journal of the American Statistical Association*, 70, (March), 113–119.

Hsu, Der-Ann, Miller, Robert B., and Wichern, Dean W. (1974) "On the Stable Paretian Behavior of Stock-Market Prices," *Journal of the American Statistical Association,* 69 (March), 108–13.

Huber, P. J. (1973). "Robust Regression: Asymptotics, Conjectures and Monte Carlo," *Annals of Statistics,* 1 (September), 799–821.

Huberman, Gur (1979). "Arbitrage Pricing Theory: A Simple Approach." Working Paper, School of Organization and Management, Yale University, December.

Jarrow, Robert, and Rudd, Andrew (1980). "Real Asset Markets and the Arbitrage Pricing Theorem." Working Paper, Graduate School of Business and Public Administration, Cornell University, February.

Jensen, Michael C. (1972). "Capital Markets: Theory and Evidence," *Bell Journal of Economics and Management Science,* 3 (Autumn), 357–98.

Kaplan, Robert S., and Roll, Richard (1972). "Investor Evaluation of Accounting Information: Some Empirical Evidence," *Journal of Business,* 45 (April), 225–57.

King, Benjamin F. (1966). "Market and Industry Factors in Stock Price Behavior," *Journal of Business,* 39 (January, Suppl.), 139–90.

Lintner, John (1965). "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics,* 47 (February), 13–37.

Mandelbrot, Benoit (1963). "The Variation of Certain Speculative Prices," *Journal of Business,* 36 (October), 394–419.

Mandelbrot, Benoit, and Taylor, H. (1967). "On the Distribution of Stock Price Differences," *Operations Research,* 15, 1057–62.

Markowitz, Harry M. (1959). *Portfolio Selection: Efficient Diversification of Investments.* New York: Wiley.

Merton, Robert C. (1973). "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41 (September), 867–87.

Mossin, Jan (1973). *Theory of Financial Markets*. Englewood Cliffs, N.J.: Prentice-Hall.

Officer, Robert R. (1972). "The Distribution of Stock Returns," *Journal of The American Statistical Association*, 67 (December), 807–12.

Ohlson, James A., and Garman, Mark B. (1978). "A Dynamic Equilibrium for the Ross Arbitrage Model." Working Paper, Department of Business Administration, University of California, Berkeley, July.

Press, S. James (1967). "A Compound Events Model for Security Prices," *Journal of Business*, 40 (July), 317–35.

Roll, Richard (1970). *The Behavior of Interest Rates*. New York: Basic Books.

Roll, Richard (1976). "Can Two-Factor Models of Asset Returns be Tested?" *European Finance Association, 1975 Proceedings*. Amsterdam: North-Holland Publ.

Roll, Richard (1977). "A Critique of the Asset Pricing Theory's Tests," *Journal of Financial Economics*, 4 (March), 129–76.

Roll, Richard, and Ross, Stephen A. (1980). "An Empirical Investigation of the Arbitrage Pricing Theory." *Journal of Finance* 35, (December), 1073–1103.

Rosenberg, Barr, and McKibben, Walt (1973). "The Prediction of Systematic and Specific Risk in Common Stocks," *Journal of Financial and Quantitative Analysis*, 8 (March), 317–33.

Rosenberg, Barr, and Houglet, Michel (1974). "Error Rates in CRSP and Compustat Data Bases and Their Implications," *Journal of Finance*, 29 (September), 1303–10.

Ross, Stephen (1976). "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 3 (December), 343–62.

Schaefer, Stephen, Brealey, Richard, Hodges, Stewart, and Thomas, Howard (1975). "Alternative Models of Systematic Risk," in Edwin J. Elton and Martin J. Gruber, eds., *International Capital Markets*. Amsterdam: North-Holland Publ.

Scholes, Myron (1972). "The Market for Securities: Substitution Versus Price Pressure and the Effects of Information on Share Prices," *Journal of Business*, 45 (April), 179–211.

Schwartz, Robert A., and Altman, Edward J. (1973). "Volatility Behavior of Industrial Stock Price Indices," *Journal of Finance*, 28 (September), 957–71.

Sharpe, William F. (1970). *Portfolio Theory and Capital Markets*. New York: McGraw-Hill.

Sharpe, William F. (1964). "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19 (September), 425–42.

Sharpe, William F. (1980) *Investments*, 2nd ed., Englewood Cliffs, N.J.: Prentice-Hall.

Watts, Ross (1973). "The Information Content of Dividends," *Journal of Business*, 46 (April), 191–211.

[illegible]