# CHAPTER IV

## EMPIRICAL TESTS OF SELECTED MODELS: SOME STATISTICAL CONSIDERATIONS

What sort of criteria must a "good" model of equity valuation satisfy? There are at least three objectives that might be specified:

(1) The variables of the model should be significant. The explanatory variables should be directly associated with the underlying causal process instead of being surrogate values of the dependent variable.

(2) The estimated parameters should exhibit reasonable stability in different cross-section samples unless there are obvious market imperfections or anomalies that have not been incorporated into the model.

(3) The parameters should exhibit reasonable stability over time. If the parameters are not stable it should at least be possible to infer some of the events causing the parameter changes.

The statistical problems associated with the testing of equity valuation models are not inherently different from the statistical complications that arise in testing most economic models. The problem is somewhat more acute in equity valuation models, however, for the parameter results must often be exact to have a useful theory. Whereas in the tests of some hypotheses it is sufficien to know parameter signs or orders of magnitude, in equity valuation models it is often necessary to have precise values if the results are to be useful to investors or firms for decision purposes.

In the remainder of this chapter we shall examine some of the statistical problems common to all of the models tested. These include the standard problems associated with least-squares regressions such as independence of the error terms, multicollinearity, errors in the variables.[1] Also discussed are the covariance tests used to test the stability of the parameter estimates and the validity of the assumptions underlying these tests.[2] We begin with a discussion of the data samples and some of their basic characteristics.

## 4.1 The Test Samples

In order to test the models it was necessary to select a variety of samples. Four sample groups of firms were selected and each group was tested for each of the four years 1956-1959. Thus, as Table IV-1 indicates, there were sixteen basic regressions for each of the models tested.

TABLE IV-1

BASIC REGRESSIONS PERFORMED ON EACH MODEL

| Sample Group | 1956 | 1957 | 1958 | 1959 |
|---|---|---|---|---|
| I | 1 | 2 | 3 | 4 |
| II | 5 | 6 | 7 | 8 |
| III | 9 | 10 | 11 | 12 |
| IV | 13 | 14 | 15 | 16 |

---

[1]For a discussion of these problems see J. Johnston, Econometric Methods, (New York: McGraw-Hill, 1963). Comments applicable to equity valuation models can be found in the first section of the piece by Irwin Friend and Marshall Puckett, "Dividends and Stock Prices," American Economic Review, 54 (September, 1964), pp. 656-662.

[2]For a discussion of this test procedure see Johnston, op. cit., pp. 136-137 or Gregory Chow, "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," Econometrica, 28 (July, 1960), pp. 591-605.

The firms selected are not random samples, for the data are derived from the firm financial statistics provided on the industrial tapes of Standard and Poor Corporation's Compustat Service.[3] The firms are all major industrial companies listed on the New York Stock Exchange. All firms for which there were missing data for the years 1951-1959 were excluded. Almost all the firms had positive earnings and dividends for the period under consideration. Three of the samples can be loosely characterized as industry type samples while the fourth sample is one that was selected from many industries. The samples include (1) a food-related industry sample of 55 firms (Group I), (2) a group of 62 firms from machinery-related industries (Group II), (3) a chemical process-related sample of 50 firms (Group III), (4) a sample of 55 of the largest firms from about fifty different industries (Group IV).[4] The first three groups do not have any common firms, but Group IV contains 22 firms that are also included in the first three groups.

The basic source data matrix contained up to 40 items of information (data points) per firm per year for each of the years 1947-1962. This aggregates to something over 140,000 data points. With this quantity of data, there must be some concern with data validity. The problem was

---

[3] The data are a preliminary 1963 version of the Compustat industrial tape made available by Standard and Poor Corporation and Bankers Trust Company of New York. For a description of the data see Compustat Users Manual, (New York: Standard and Poor Corporation, 1966) or IBM Financial Analysis Program, (White Plains, N.Y.: IBM Technical Publications, 1964).

[4] The industry designations that have been assigned to companies in Compustat are made by the four digit Standard and Poor code. This code closely follows the U. S. Government Standard Industry Classification code although there have been some additions and some switching of firms.

further complicated in this research project by the fact that the data were processed by three noncompatible computer centers. The data for about 25% of the firms have been spot-checked by students as a classroom exercise. Each time the data were transferred from one computer center to another it was necessary to punch the data onto cards from the old computer center tapes and reload it on the tapes of the new computer center. During the reloading order of magnitude and sign checks were performed on most of the data. This check procedure detected approximately 20 cards punched with random bits out of the 24,000 cards processed. The tapes are checked periodically by re-running some of the regression programs to see that the results have not changed.

The years 1956-1959 were chosen as the test period both because the data for this period were relatively complete and because this period contained a significant stock market cycle. At the end of 1956 stock prices, as measured by Standard and Poor's Industrial Stock Price Index, had been stable for almost a year. By the end of 1957, the index had undergone a substantial decline, but a snapshot one year later of year-end 1958 prices showed relative positions after the index had been rising steadily for 12 months. The year 1959 was another one of relative stability with the index up moderately at the end of the year. Table IV-2 summarizes the price, earnings, dividends figures for the Standard and Poor Index and for each of the four samples. The Standard and Poor averages are weighted averages with the weighting factor being relative aggregate market value of the common stock equity, while the sample averages are just simple equal-weight averages. Note that the price averages

TABLE IV-2

AVERAGE VALUES FOR PRICE, NET INCOME, DIVIDENDS

| Group | | 1956 | 1957 | 1958 | 1959 |
|---|---|---|---|---|---|
| S. &. P. | P | 50.08 | 42.86 | 58.97 | 64.50 |
| Industrial Index | ni | 3.53 | 3.50 | 2.95 | 3.53 |
| | dv | 1.78 | 1.84 | 1.79 | 1.90 |
| | P | 31.07 | 30.96 | 42.83 | 47.01 |
| I | ni | 3.32 | 3.42 | 3.55 | 3.76 |
| | dv | 1.73 | 1.79 | 1.85 | 1.96 |
| | P | 39.19 | 27.20 | 38.56 | 43.72 |
| II | ni | 3.86 | 3.41 | 2.21 | 3.33 |
| | dv | 1.92 | 1.91 | 1.57 | 1.62 |
| | P | 46.22 | 41.32 | 56.31 | 63.25 |
| III | ni | 3.19 | 3.15 | 2.70 | 3.28 |
| | dv | 1.58 | 1.68 | 1.62 | 1.74 |
| | P | 47.39 | 41.03 | 58.01 | 67.10 |
| IV | ni | 3.91 | 3.64 | 3.06 | 3.65 |
| | dv | 1.94 | 1.96 | 1.91 | 2.01 |

Standard and Pool index values are from Standard and Poor's Trade and Securities Statistics, Vol. 30, No. 4 (1964), pp. 123-124. All figures are on an adjusted per share basis.

of all the groups fell sharply at the end of 1957 and recovered strongly in 1958 despite generally lower earnings and dividends. This type of movement is likely to lead to rapidly changing parameter levels in equilibrium models that are restricted to firm financial variables.

None of the samples are in any sense perfect "pure industry samples" or "homogeneous risk class" samples. This should not be so important in the Benishay or Gordon models where risk measures are supposedly incorporated as explicit measures in the models. The absence of sharp risk stratification could lead to inaccurate estimates of the parameters in the Durand, Modigliani-Miller, and Barges models however. Unfortunately none of these authors adequately specifies a procedure for selecting a homogeneous risk class sample. It simply will not do, of course, to say a posteriori that the parameter estimates are not satisfactory because the sample was not homogeneous in an important omitted variable. If the concept "risk class" is to be meaningful, some a priori screening procedure must be specified.

One might conjecture that a homogeneous risk class could be constructed from firms (1) facing the same product markets, (2) having similar production processes, (3) having access to the same factor markets, (4) using uniform accounting procedures. The increasing complexity and product diversification of large American firms would seem to preclude the possibility of forming such pure samples.[5] An attempt was made, however, to give each of the samples some sort of homogeneity. Thus the food-related industry sample is generally characterized by firms with (a) large positive and stable earnings and dividends and (b) relatively low price-earnings ratios. The machinery-related firms sample consists of firms primarily manufacturing industrial goods. Earnings of these firms tend to be cyclical and dividends are somewhat more unstable than in the case of the Group I firms. Many of the Group III "chemical activity" firms are characterized as growth firms. The price-earnings ratios of these firms

---

[5]Two industries that do have relatively uniform accounting procedures and similar production processes across firms are the banking industry and the electric utility industry. Unfortunately, Durand ("Bank Stock Prices and the Analysis of Covariance," op. cit.) did not find homogeneity in the banking industry. Professor Miller has stated on several occasions that he feels the electric utility industry is the only one that has the necessary data accuracies for testing the M-M model. If this be true, it would severely impair the utility of the M-M model.

tend to be somewhat higher than the same ratios in the other groups. This group contains at least three identifiable product industries--basic chemicals, drugs, and petroleum. Even so, only fifty firms that had the necessary data were available. The fourth group is made up of industry leaders from about fifty different industries. Each of these firms is usually one of the largest or best known in its particular industry. This grouping was tried because there may be some possibility that the increasing diversification within American firms makes risk more a function of relative firm size than traditional industrial classification.[6] All these are still rather arbitrary attempts to construct homogeneous samples. If risk class is an important variable, the best procedure would be to include it as an explicit variable. Unfortunately, this turns out to be an extremely complicated task.

## 4.2 Standard Regression Problems[7]

The standard regression model proceeds from the following set of initial assumptions:[8]

---

[6]Firm size measures such as sales, total assets, total market value, may be surrogates for diversification of product groups, production processes, management personnel, or financial holdings. To the extent that increased size implies increased public awareness and hence increased stockholder interest, there may also be risk reduction due to a broader financial market for a firm's stock.

[7]Sections 4.2-4.5 have benefited considerably from the exemplary discussion by Edwin Kuh in his description of the statistical problems associated with testing investment theory models. See Edwin Kuh, Capital Stock Growth: A Micro-Econometric Approach, (Amsterdam, Netherlands: North-Holland Publishing, 1963).

[8]Reproduced from Kuh, op. cit., pp. 87-88 and Johnston, op. cit., p. 107.

(1) The dependent variable is the sum of a linear function of the explanatory variables and an error term with constant variance.

(2) The explanatory variables are usually assumed fixed in the sense that no probabilistic mechanism is postulated for generating them. The explanatory variables can be random variables, but they must be independent of each other. Further, the error term is uncorrelated with any of the explanatory variables.

(3) The samples are independent. That is, the error terms are independently distributed random variables. To derive significance tests for parameter estimates it is assumed that the error terms are normally distributed.

Estimated parameters may be biased or subject to undue variation because of a variety of problems.

(1) Errors in variables: Most of the models discussed deal with expectational variables. The actual measures of price, net income, and growth may not be adequate surrogates for these expectational variables. Since the problem would appear to be particularly acute for expected net income, some of the parameters associated with this variable may be underestimates of their true values.[9]

(2) Autocorrelation: Omitted variables, incorrect model specification, or other factors could lead to autocorrelation in the disturbance terms. This could result in an underestimation of the regression sampling variances.

(3) Multicollinearity: Since reasonably sharp parameter estimates are desired, excessive multicollinearity between explanatory variables could be a problem. As it turned out, the only independent variables that were even moderately highly correlated were earnings and dividends. Table IV-3 summarizes the simple correlations between prices, earnings, and dividends for the various samples. This combination of variables was used only in the Durand models.

(4) Omitted variables: It is obvious from the varying explanatory powers of the different models that significant variables have sometimes been omitted. All the models restrict themselves to historical firm financial variables. Thus, at least two potential classes of variables are ignored: (a) market-interaction, investor psychological variables, (b) relative macro-economic (industry or national economy) measures.

---

[9] See Johnston, op. cit., pp. 148-175.

TABLE IV-3

SIMPLE CORRELATION COEFFICIENTS

| Group | | 1956 P | 1956 ni | 1956 dv | 1957 P | 1957 ni | 1957 dv | 1958 P | 1958 ni | 1958 dv | 1959 P | 1959 ni | 1959 dv |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P | 1.00 | .80 | .93 | 1.00 | .88 | .94 | 1.00 | .86 | .91 | 1.00 | .88 | .89 |
| I | ni | .89 | 1.00 | .83 | .90 | 1.00 | .87 | .86 | 1.00 | .86 | .89 | 1.00 | .83 |
| | dv | .92 | .82 | 1.00 | .93 | .85 | 1.00 | .93 | .86 | 1.00 | .89 | .87 | 1.00 |
| | P | 1.00 | .78 | .62 | 1.00 | .69 | .72 | 1.00 | .69 | .76 | 1.00 | .60 | .66 |
| II | ni | .82 | 1.00 | .86 | .55 | 1.00 | .75 | .68 | 1.00 | .66 | .68 | 1.00 | .55 |
| | dv | .62 | .63 | 1.00 | .71 | .46 | 1.00 | .62 | .55 | 1.00 | .55 | .64 | 1.00 |
| | P | 1.00 | .74 | .78 | 1.00 | .69 | .78 | 1.00 | .71 | .78 | 1.00 | .78 | .79 |
| III | ni | .72 | 1.00 | .78 | .71 | 1.00 | .69 | .74 | 1.00 | .73 | .68 | 1.00 | .83 |
| | dv | .61 | .56 | 1.00 | .67 | .51 | 1.00 | .76 | .74 | 1.00 | .66 | .61 | 1.00 |
| | P | 1.00 | .63 | .49 | 1.00 | .63 | .60 | 1.00 | .63 | .52 | 1.00 | .62 | .49 |
| IV | ni | .74 | 1.00 | .76 | .68 | 1.00 | .76 | .55 | 1.00 | .74 | .48 | 1.00 | .83 |
| | dv | .67 | .64 | 1.00 | .74 | .72 | 1.00 | .74 | .66 | 1.00 | .53 | .89 | 1.00 |

Note: correlations above the diagonals are between the variables price (P), per share net income after taxes (ni), and per share dividends (dv); correlations below the diagonals are between the logarithms of these variables.

(4) Omitted variables: (Continued)
Unless something is known about the correlations between these omitted variables and variables in the models, nothing can be said about the types of biases such omissions might cause.

(5) Least-squares estimations: Least-square estimating procedures are such that an individual data point is more heavily weighted the more it departs from the sample average. Thus, regression coefficients can be substantially influenced by extreme values. Unfortunately, there is evidence to suggest that present models of equity valuation are poor theoretical bases from which to explain prices given extreme value data points.[10]

(6) Mis-specification of model type: All of the present models are assumed to be linear or log-linear relations with simple non-constrained independent variables. If this is the wrong model type, the sharp results desired might never be achieved, no matter how many new variables are tried.

To test the assumption of a constant variance the estimated variances from the sixteen basic regressions for each model were compared. First the estimated variances from all sixteen regressions were compared. Then the equality of estimated variances across time for a given group was tested. Finally the sample variances across groups for a given time period were compared. The test statistic used in testing for the equality of the variances was the $[-2 \log \lambda]$ statistic described by Mood and others.[11] The results are summarized in Table IV-4. These results strongly suggest that variances cannot be considered equal across Groups I - IV, but that variances are probably equal across years 1956-1959 for each group.

The normality of the residuals was examined by dividing the residuals for each regression for the year 1956 into cells and comparing the observed

---

[10]For example, many researchers exclude firms with negative earnings or zero dividends from their test samples. The formation of true expectations about firms with extended periods of losses, or zero dividends, or high absolute growth rates, is not adequately described by most existing empirical specifications of equity valuation models.

[11]See Alexander Mood, Introduction to the Theory of Statistics, (New York: McGraw-Hill, 1950), pp. 267-270.

TABLE IV-4

TESTS OF EQUALITY OF SAMPLING VARIANCES

| | ALL | I | II | III | IV | 1956 | 1957 | 1958 | 1959 |
|---|---|---|---|---|---|---|---|---|---|
| Linear Durand | 877.* | 20.* | 10.* | 6 | 14.* | 236.* | 193.* | 231.* | 167.* |
| Log Durand | 133.* | 5. | 3. | 1. | 5. | 29.* | 35.* | 27.* | 37.* |
| M-M | 162.* | 8. | 19.* | 9.* | 6. | 21.* | 41.* | 45.* | 32.* |
| Barges | 140.* | 4. | 18.* | 11.* | 8. | 23.* | 31.* | 40.* | 21.* |
| Benishay | 98.* | 3. | 3. | 1. | 4. | 28.* | 24.* | 27.* | 15.* |
| Gordon | 120.* | 1. | 11.* | 2. | 1. | 24.* | 31.* | 38.* | 21.* |

* = Significant at 5% level.
I = Assumed equality of variances for tests (1, 2, 3, 4).
1956 = Assumed equality of variances for tests (1, 5, 9, 13) - See Table IV-1.

frequency in each cell with expected frequency under the assumption that the errors were normally distributed. Results of the Chi-squared tests are reported in Table IV-5. An examination of selected samples for other years suggested the results would not differ materially from year to year. In only three cases was the assumption of normality rejected. Each of these cases was in Group IV and each rejection was due to the fact that IBM had a very large positive residual causing the other residuals to be bunched much closer to the mean than would be predicted.

If one is particularly concerned about departures from normality of extreme values the Chi-squared test may not be sufficient by itself. In order to make the cell count size reasonable observations from about two standard deviations out tend to be put into one cell. Recent developments

TABLE IV-5

CHI-SQUARED TEST OF NORMALITY OF RESIDUALS, 1956

| Group | I | II | III | IV |
|---|---|---|---|---|
| Linear Durand | 9.4 | 6.9 | 2.7 | 38.5* |
| Log Durand | 2.0 | 5.5 | 1.9 | 7.9 |
| M-M | 3.6 | 5.5 | 4.0 | 25.8* |
| Barges | 5.5 | 2.5 | 4.2 | 21.0* |
| Benishay | 4.2 | 6.2 | 5.9 | 2.0 |
| Gordon | 1.0 | 3.5 | 3.1 | 3.2 |

*significant at 5% level

in research on price fluctuations suggest that small departures from normality, particularly in extreme values, can lead to radically different expectations about the variance of the residuals for the distribution under study.[12] For the twenty-four 1956 regressions there were nine cases where the residual was greater than four standard deviations from the mean. Six of these cases were the IBM residual. If Group IV is ignored (the group that contained IBM), that still leaves three cases out of eighteen regressions to consider. The expected frequency for observations greater than four standard deviations from the mean in a normal distribution is about two in 10,000 observations. This is much lower than the observed frequency which works out for the eighteen regressions to be three in 1002 observations.

---

[12]See the discussion on the properties of the stable Paretian distribution by Eugene Fama, "The Behavior of Stock-Market Prices," Journal of Business, 38 (January, 1965), pp. 34-106.

One tends to feel, after examining residuals from a large number of regressions in the equity valuation area, that extreme values are considerably more frequent than would be warranted under assumptions of normality. Resources and data limitations do not permit a test of the hypothesis that equity valuation model regression residuals follow a stable Paretian distribution rather than a Gaussian distribution. That hypothesis would be subject matter enough for more than one dissertation.

## 4.3 Firm Effects on Sample Residuals

The four samples described in section (4.1) are not completely independent since Group IV contains some firms found in the other three groups. Still the firms could be statistically independent in the sense that the error terms appear as independently distributed random variables over time for the same firm. To test for what Kuh has called "firm effects" a signs test was constructed for each of the samples.[13] The null hypothesis is that within a given sample (Groups I, II, III, IV) the signs of the residuals for the same firm are randomly distributed over time (years 1956-1959) with equal probability of being positive or negative.

Positive deviation signs were counted for each firm. Since there were four years, the measure could assume the value [0, 1, 2, 3, 4]. The frequency count for each value was tabulated for the sample as a whole and compared to the expected frequency distribution.[14] Table IV-6

---

[13] See E. Kuh, op. cit., pp. 158-188.

[14] That is, the number of firms in the sample that had zero positive residuals was tabulated, the number of firms with one positive residual, etc.

TABLE IV-6

TEST OF FIRM RESIDUAL INDEPENDENCE OVER TIME

| Number of (+) Dev. | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| GROUP I | | | | | |
| (Expected Frequencies) | 4 | 13 | 21 | 13 | 4 |
| Linear Durand | 20 | 5 | 7 | 8 | 15 |
| Log Durand | 17 | 5 | 6 | 8 | 19 |
| M-M | 16 | 5 | 14 | 8 | 12 |
| Barges | 18 | 9 | 9 | 9 | 10 |
| Benishay | 20 | 8 | 5 | 6 | 16 |
| Gordon | 14 | 10 | 8 | 10 | 13 |
| GROUP II | | | | | |
| (Expected Frequencies) | 4 | 15 | 24 | 15 | 4 |
| Linear Durand | 19 | 12 | 11 | 12 | 8 |
| Log Durand | 12 | 14 | 16 | 8 | 12 |
| M-M | 6 | 17 | 14 | 15 | 10 |
| Barges | 8 | 13 | 20 | 12 | 9 |
| Benishay | 18 | 11 | 5 | 9 | 19 |
| Gordon | 13 | 15 | 12 | 14 | 8 |
| GROUP III | | | | | |
| (Expected Frequencies) | 3 | 12 | 20 | 12 | 3 |
| Linear Durand | 15 | 8 | 5 | 11 | 11 |
| Log Durand | 11 | 10 | 10 | 7 | 12 |
| M-M | 16 | 7 | 10 | 6 | 11 |
| Barges | 15 | 10 | 9 | 6 | 10 |
| Benishay | 19 | 6 | 4 | 3 | 18 |
| Gordon | 15 | 11 | 3 | 4 | 17 |
| GROUP IV | | | | | |
| (Expected Frequencies) | 4 | 13 | 21 | 13 | 4 |
| Linear Durand | 9 | 23 | 8 | 7 | 8 |
| Log Durand | 19 | 4 | 8 | 10 | 14 |
| M-M | 17 | 13 | 10 | 9 | 6 |
| Barges | 21 | 4 | 13 | 9 | 8 |
| Benishay | 14 | 9 | 6 | 12 | 14 |
| Gordon | 20 | 11 | 1 | 7 | 16 |

All Chi-square tests significant at 5% level.

indicates the expected frequency count and actual count for each sample for each of the six models tested. Using a Chi-squared test all differences between actual and expected frequencies were found to be significant at the 5% level. In fact, in fifteen of the twenty-four regression groups the majority of firms had residual signs that were either all positive or all negative.

It is interesting to note that the firm effects are present to about the same degree in all six regression models. That is, adding additional variables or changing from a linear to logarithmic specification did not seem to reduce these firm effects. Of course, firm effects are important only if data are to be pooled across time. But that is exactly what one would like to do in order to evaluate temporal parameter stability.

## 4.4 Industry Effects on Sample Residuals

Omitted variables, incorrect model specification, or other factors can lead to autocorrelation in the disturbance terms. Such autocorrelation can result in serious underestimates of the sampling variances.[15] While the regression slope parameters are unbiased, the sampling variances may be unduly large compared with those achievable by alternative methods to straightforward least-squares procedures. Since the effective number of degrees of freedom may be considerably less in an autocorrelated structure than the nominal number of degrees, any F tests will be biased downward and result in rejection of the F ratio hypothesis (no significant difference between numerator and denominator variances) more often than is warranted.

---

[15] For a discussion of these problems see Johnston, op. cit., pp. 177-200.

To test for autocorrelation generated by a single Markov process, Durbin-Watson (d) statistics were computed for each of the ninety-six basic regressions.[16] Normally, of course, one would not expect this type of autocorrelation in cross-section samples. But the tests were made, and their significance led to some interesting findings about unexplained industry effects. The results tabulated in Table IV-7, indicate a

TABLE IV-7

DURBIN-WATSON TESTS OF BASIC REGRESSIONS

| | | 1956 | 1957 | 1958 | 1959 |
|---|---|---|---|---|---|
| Log Durand | I | 1.33* | 1.27* | 1.03* | 1.61 |
| | II | 1.64 | 1.83 | 2.05 | 2.07 |
| | III | 1.05* | 1.19* | 1.54 | 1.22* |
| | IV | 1.93 | 2.04 | 1.92 | 1.76 |
| Linear Durand | I | 1.54 | 1.36* | 1.06* | 1.51 |
| | II | 1.89 | 1.88 | 1.93 | 2.20 |
| | III | 1.25* | 1.28* | 1.16* | 1.26* |
| | IV | 1.97 | 1.92 | 1.95 | 1.74 |
| Modigliani-Miller | I | 1.48 | 1.47 | 1.79 | 1.77 |
| | II | 1.57 | 1.53 | 1.64 | 1.79 |
| | III | 1.49 | 1.44 | 1.49 | 1.39* |
| | IV | 2.01 | 1.94 | 1.72 | 1.44 |
| Barges | I | 1.43 | 1.36* | 1.69 | 1.66 |
| | II | 1.61 | 1.56 | 1.65 | 1.68 |
| | III | 1.37* | 1.31* | 1.47 | 1.35* |
| | IV | 2.00 | 1.86 | 1.76 | 1.48 |
| Benishay | I | 1.73 | 1.87 | 1.59 | 1.57 |
| | II | 2.03 | 2.02 | 1.81 | 2.07 |
| | III | 1.26* | 1.53 | 1.26* | 1.32 |
| | IV | 1.93 | 2.47 | 1.96 | 1.90 |
| Gordon | I | 1.93 | 1.81 | 1.70 | 1.98 |
| | II | 1.85 | 1.79 | 1.88 | 2.43 |
| | III | 1.35 | 1.30 | 1.41 | 1.02* |
| | IV | 1.83 | 1.80 | 1.90 | 1.63 |

*significant at 5% level.

[16] That is, we are testing for first order correlations of the error terms to see if a relation of the form ($E_x = \alpha E_{x-1} + v_x$) exists, where (x) is the firm position in the sample and (v) is a random variable independently distributed.

significant positive serial correlation at the 5% level in twenty cases.[17] This is obviously too large a number of cases to be considered random, so a further investigation was made to see why these cross-section regressions had such significant autocorrelations in their residuals.

An examination of the signs of the residuals revealed some extraordinarily long runs of positives or negatives for each sample group. Moreover these runs did not seem to vary much from year to year or even regression model to regression model. When the sign sequences were compared to the lists of firms, it became clear that some sort of industry effect was present.

To test for this industry effect each sample group was divided into four cells. The number of firms in each cell was dependent on both a "natural industry break" and a change in residual signs as indicated by the linear Durand results. Thus for Group I we have four cells made up of (1) large cereal, meat, and milk companies, (2) baking and sugar companies, (3) distillers and soft drink companies, and (4) tobacco products industries. For Group II, the cells were (1) brass and copper processors, (2) form equipment manufacturers, (3) large machine tool companies, (4) specialty machine tool companies. Group III was broken down into (1) chemical companies, (2) drug companies, (3) domestic oil companies, (4) selected international oils. The hypothesis is that autocorrelation occurred because industry effects within each sample

---

[17] Significance levels estimated from Table 5 in the article by J. Durbin and G. Watson, "Testing for Serial Correlation in Least Squares Regression II," Biometrika, 1951, pp. 159-176. See also the analysis by H. Theil and A. Nager, "Testing the Independence of Regression Disturbances," Journal of American Statistical Association, 56 (1961), pp. 793-806.

were present. Since firms are arranged sequentially on the Compustat tapes according to industry classification, the unspecified industry effects would create clustering and tend to produce residuals of the same sign for a given "industry." It must be noted here that the cell groupings do not strictly conform to traditional industry groups but are somewhat broader. Furthermore, since the linear Durand residuals were used to help determine breaking points, the cell divisions are probably not those that would have been selected on an *a priori* basis.

To test the hypothesis of industry effects for a sample group, the number of positive residuals over all four years for firms in a cell was tabulated. Thus if there were fifteen firms in a cell, there could be up to (15 * 4) sixty positive residuals. The number of positive residuals counted was compared to the expected number assuming a random equiprobable distribution of residual signs. One would expect the computed Chi-squared statistic to be significant for those cases in the first three groups where the Durbin-Watson statistic was significant. The Chi-squared statistic should not be significant in Group IV since firms in this group come from many industries with only very short sequences from the same general "industry" area.

Table IV-8 indicates that the expected results were fully realized. Indeed, the Chi-squared statistics were significant in every instance in sample groups I-III and in no instance for sample group IV. Moreover, the orders of magnitude of the dominant sign were reasonably stable from cell to cell across regressions, suggesting that the industry effects

TABLE IV-8

TEST OF RESIDUALS FOR POSSIBLE INDUSTRY EFFECTS

(Number of Positive Residuals Over a Four Year Period)

| | | | | |
|---|---|---|---|---|
| GROUP I: firms | 1-14 | 15-36 | 37-46 | 47-55 |
| (Expected Values) | 28 | 44 | 20 | 18 |
| Linear Durand | 47 | 18 | 20 | 14 |
| Log Durand | 51 | 26 | 26 | 14 |
| M-M | 15 | 61 | 11 | 18 |
| Barges | 12 | 60 | 12 | 10 |
| Benishay | 12 | 54 | 13 | 21 |
| Gordon | 46 | 29 | 17 | 16 |
| GROUP II: firms | 1-10 | 11-20 | 21-34 | 35-62 |
| (Expected Values) | 20 | 20 | 28 | 56 |
| Linear Durand | 12 | 27 | 31 | 32 |
| Log Durand | 13 | 25 | 37 | 43 |
| M-M | 25 | 10 | 20 | 75 |
| Barges | 23 | 10 | 21 | 70 |
| Benishay | 27 | 14 | 13 | 67 |
| Gordon | 11 | 24 | 33 | 44 |
| GROUP III: firms | 1-17 | 18-36 | 37-46 | 47-50 |
| (Expected Values) | 34 | 38 | 20 | 8 |
| Linear Durand | 57 | 14 | 26 | 0 |
| Log Durand | 53 | 16 | 23 | 7 |
| M-M | 9 | 45 | 21 | 14 |
| Barges | 9 | 43 | 20 | 14 |
| Benishay | 13 | 35 | 31 | 16 |
| Gordon | 54 | 28 | 15 | 0 |
| GROUP IV: firms | 1-14 | 15-28 | 29-42 | 43-55 |
| (Expected Values) | 28 | 28 | 28 | 26 |
| Linear Durand | 25 | 24 | 25 | 18 |
| Log Durand | 25 | 30 | 27 | 24 |
| M-M | 24 | 16 | 23 | 22 |
| Barges | 24 | 16 | 21 | 28 |
| Benishay | 28 | 31 | 29 | 25 |
| Gordon | 24 | 28 | 30 | 16 |

Chi-squared tests for Groups I-III all significant at 5%.
Firm numbers - names are found in Appendix Tables 17-20, Chapter V.

did not disappear for any of the different regression frameworks.[18]

While the significance of the test statistics for groups I-III in Table IV-8 does suggest a plausible reason for the significance of so many Durbin-Watson statistics, Table IV-8 does not imply that autocorrelation or "industry effects" are a serious problem in every instance. Table IV-8 is merely a signs test, and the Durbin-Watson test for autocorrelation a magnitude test. To the extent that the Durbin-Watson test accurately measures autocorrelation, then the problem exists primarily with the Durand regressions where the absence of any growth or risk variables undoubtedly accentuates apparent industry effects.

## 4.5 Covariance Tests of Parameter Stability

All of the equity valuation models studied in this work and virtually all equity valuation models for which tests have been reported in the literature "confirm" the validity of the estimates made by testing the same sample for a number of years or a cross-section of samples for the same year. Unfortunately the discussions of parameter consistency across samples or stability over time tend to be qualitative and impressionistic. Quantitative studies of parameter stability are virtually absent from the literature on equity valuation models.

None of the authors reviewed suggest that the standard assumptions of regression analysis are inappropriate for their model tests. Nor do

---

[18] Note that in some regressions the signs for an industry were positive while in other regression models the same signs were negative. To compare orders of magnitude convert the cells where negatives predominated to positive equivalents by subtracting the cell count from (2* expected value). Then the counts in all cells will be ($\geq$) expected value and ($\leq$) maximum cell count possible and order of magnitude comparison will be possible.

these authors suggest that there are omitted variables which would systematically alter parameter estimates across samples or over time.[19] But if these conditions hold, there are tests available which will allow the researcher to ascertain whether or not different samples could come from the same underlying population.

The tests that will be used in this work to study parameter stability are analysis of covariance tests of equality between sets of coefficients in linear regressions. The null hypothesis is that parameter estimates for the selected samples are all from the same underlying population.[20] That is,

$$H_o: \beta_1 = \beta_2 = ---\beta_k = \beta$$

where ($\beta$) is the vector of coefficients for the specified relation. The analysis of covariance tests essentially compare the residuals from the sample regressions assumed to come from the same population with the residuals from a pooled regression made up of all the sample data under consideration. The types of tests performed are indicated in Table IV-9. The actual test results will be reported in Chapter V.

Since it is obviously desirable to use the maximum sample available to increase the sharpness of parameter estimates, the first covariance test was a comparison of all sixteen regressions for a model to see if they came from the same population. It might very well be that all the

---

[19] The exception to this statement is the Modigliani-Miller comments on risk class. Their discussions, however, are qualitative and essentially suggest that the experimenter himself must somehow perform a screening activity the model itself cannot perform.

[20] For a discussion of these analysis of covariance procedures, see the references in footnote (2) and also Kuh, op. cit., pp. 127-148.

TABLE IV-9

SETUP OF HYPOTHESES FOR ANALYSIS OF COVARIANCE TESTS

| Test | Parameters of These Regressions Were Assumed Equal (See Table IV-1) |
| --- | --- |
| All | All Sixteen Regressions |
| I | 1 - 2 - 3 - 4 |
| II | 5 - 6 - 7 - 8 |
| III | 9 - 10 - 11 - 12 |
| IV | 13 - 14 - 15 - 16 |
| 1956 | 1 - 5 - 9 - 13 |
| 1957 | 2 - 6 - 10 - 14 |
| 1958 | 3 - 7 - 11 - 15 |
| 1959 | 4 - 8 - 12 - 16 |

data could not be pooled, but data for the same sample group could be pooled for all years or data for the same year could be pooled across sample groups. Therefore, if the "all data" hypothesis was rejected the regressions were grouped first by sample group and then by year to see if any of these sample groupings had parameter estimates that could have come from the same population. Tests 2-5 in Table IV-9 are the tests that hold the sample index constant and assume the equality of results across years. It is somewhat surprising to find so many willing to compare parameter estimates for different years since none of the models correct for macro-economic fluctuations, but such comparisons do seem to be standard practice. The last four tests in Table IV-9 hold the year index constant and assume the equality of results across samples. This should be true only to the extent there are not significant risk variations from sample to sample.

It must be noted that several of the statistical prerequisites for making analysis of covariance tests are not strictly fulfilled for models in this study. First, as indicated in section 4.2, there is some uncertainty as to the normality of the residuals. Second, it is assumed that the data under consideration all have equivalent variances. But as Table IV-4 indicates, this condition is not fulfilled for the cross-section sampling variances, although the condition is fulfilled for the across-years sampling variances for each group. It is not clear what impact departures from error variance equality will have on the ultimate F tests. Finally, the procedure outlined involves some double sampling.[21] Unfortunately, known tests of significance cannot easily handle such a situation. However, there are precedents for making such tests and assuming available tests of significance are approximately correct.[22] Despite the indicated problems, since it is crucial that we know something about parameter stability, the covariance tests were made.

In this chapter we have examined a variety of problems associated with the assumptions of standard regression theory. The statistics presented raise serious questions about the validity of such assumptions

---

[21]For example, regression residuals (1) appear in the (F) tests for tests (ALL), (I), and (1956).

[22]See Kuh, op. cit., pp. 127-133. Kuh's tests, however, are more nearly a logical sequence than the tests outlined here. That is, his (F) tests cascade so that making the second test depends on the results of the first test, etc., where virtually the same data are used in some of the sequential tests. The tests in Table IV-9 are logically independent and the data only partially overlap.

for the models surveyed. In the next chapter the parameter estimates and covariance tests are presented. There, too, we shall see results that must raise questions about the appropriateness of existing models of equity valuation.