Division of Research
Graduate School of Business Administration
The University of Michigan

August 1976

METHODOLOGICAL PROBLEMS WITH PRICE-EARNINGS RATIO ANALYSIS

Working Paper No. 134

by

Joseph K. Cheung

The University of Michigan

Introduction

In recent years, ordinary least squares regression techniques (OLS) have been widely used by researchers in accounting and related areas.[1] Under a certain set of conditions, OLS is an appropriate statistical model for summarizing relationships among data under investigation.

Because OLS is a cross-sectional analysis, the researcher must be willing to accept the regression coefficients as estimates of the parameter values in equilibrium. Furthermore, the researcher who uses OLS necessarily assumes that his data are well-behaved, i.e., there are no extreme outliers and the variables are symmetrically distributed. If the researcher cannot be comfortable with these conditions, then he should have reservations about the appropriateness of OLS.

The purpose of this study is to expose the weakness of OLS when these conditions cannot be entertained. This paper, the first of two, deals with the possibility of rendering analysis more reliable by pooling time-series and cross-sectional observations when such data are available. The second paper will deal with the question of outliers and the possibility of using robust estimation techniques instead of OLS.[2]

For expository purposes, we will discuss the methodological problems with OLS in the context of determining the explantory factors for price-earnings ratios (p/e) of common stocks.

Problems with Past Studies

Although investment analysts generally have regarded p/e ratios as useful valuation tools,[3] researchers have repeatedly failed to use OLS to determine the explantory factors. The major works in this area have been summarized and critically reviewed by Keenan [8]. According to Keenan, the general results are that:

> ...the greater majority of the estimated parameters for the variables of these models are neither statistically significant nor stable. About all one can conclude from existing evidence is that there is usually (for screened samples) a positive relationship of unspecified magnitude between equity share prices and dividends, earnings, and growth rates. The performance of other variables, especially risk-type variables (as measured by firm financial data), is poor as parameter signs are often indeterminant and magnitudes highly unstable. [8; p. 243-4]

In sum, the literature on the topic abounds with attempts to regress p/e ratios (or modifications thereof) on a number of variables specified differently by each researcher. Rather than providing the reader with a consistent set of significant results, each researcher develops his own set of variables. No meaningful comparisons between sets can be made.

I believe the forces that determine the magnitudes of p/e ratios are subject to different influences in different time periods. Therefore, it would be more meaningful to capture the dynamic effects over time (which can be viewed as short-run adjustments) in the cross-sectional analysis (which can be regarded as long-run equilibrium conditions). This is the thrust of the argument presented here. The statistical technique appropriate for combining both effects is a pooled time-series and cross-sectional model. This paper presents results using OLS and two different pooled models. The experimental design consists of testing for sample sensitivity and variation of parameter estimates under each technique.

Method

There are a number of techniques available in econometric and statistical literature for analyzing a time series of cross-sectional samples. Johnson and Lyon [7] have assembled experimental evidence comparing these different techniques. Their conclusion seemed to favor the covariance model originated by Suits and expounded by Schipper [11]. This study uses both the covariance model and a "cross-sectionally heteroskedastic and time-wise autoregressive model."

Covariance regression model

Covariance regression differs from ordinary least squares in that it accounts for the specific characteristics of individual units and the exogeneous influences in each time period, in addition to the explanatory variables included in the regression.

If we have data on J number of sample units for T number of time periods, the typical linear relation for the j^{th} unit at time t can be described as

$$Y_{jt} = \alpha + \beta_1 X_{1jt} + \beta_2 X_{2jt} + \dots + \beta_K X_{Kjt} + b_j + C_t \qquad (1)$$

$$(j = 1, \dots, J \; ; \; t = 1, \dots, T)$$

where y_{jt} = the dependent variable of the j^{th} unit at time t;

X_{ijt} = the i^{th} explanatory variable of the j^{th} unit at time t;

β_i = slope coefficients for the i^{th} explanatory variable;

α, b_j, c_t = parameters.

With the exception of b_j and c_t, the equation is analagous to a multiple regression model.

Cross-sectional data come from different sample units, each possessing certain characteristics that cannot be specified. A postulated cross-sectional model will never be able to fully specify all explanatory variables. Therefore, for each unit j, b_j is added to the model to capture the unspecfied factors whose effect on the dependent variable may nonetheless be important and also vary between sample units. On the other hand, analyzing a sample unit's behavior over time enables one to examine only the dynamic behavior pertaining to that sample unit but will not account for changes in the environment. To account for these changes, c_t is added to the model. Equation (1) can be interpreted so that the value of Y for the j^{th} unit at time t depends not only on the values of the independent variables X_i, but also on two parameters: b_jk which is peculiar to the j^{th} unit for all periods, and c_t, which is specific to the t^{th} period but common to all units.

The advantages of the covariance regression model have been fully described by Schipper [11]. In sum, the method attempts to enrich the analysis by utilizing more information, and yet controlling for all other factors contained in the observations on the dependent variable but not accounted for in the postulated regression equation. By doing so, the intent is to attain a higher degree of precision and reliability in estimation.

The computational techniques require several transformations of variables. If we consider the average values of the variables of the j^{th} over all time points, Equation (1) can be written as:

$$\bar{Y}_j = \alpha + \beta_1 \bar{X}_{1j} + \beta_2 \bar{X}_{2j} + \ldots + \beta_K \bar{X}_{Kj} + b_j + \bar{c}_t \quad (2)$$

where the bar symbol designates averages and the c_t is assumed to average to zero over time.

Similarly, we can compute sample averages of the variables at each time point t. This results in

$$\bar{Y}_t = \alpha = \beta_1 \bar{X}_{1t} + \beta_2 \bar{X}_{2t} + \ldots + \beta_K \bar{X}_{Kt} + \bar{b}_j + c_t \quad (3)$$

where, assuming we have a random sample, the b_j's across the sample at any time t will also average to zero.

Finally, on taking the grand mean of each variable over time and across sample units, we obtain

$$\bar{\bar{Y}} = \alpha + \beta_1 \bar{\bar{X}}_1 + \beta_2 \bar{\bar{X}}_2 + \ldots + \beta_K \bar{\bar{X}}_K + \bar{\bar{b}}_j + \bar{\bar{c}}_t \quad (\bar{\bar{b}}_j = \bar{\bar{c}}_t = 0) \quad (4)$$

where double bars represent grand means of each variable. The estimating equation can be obtained by subtracting Equations (2) and (3) from Equation (1) and then scaling the result by adding Equation (4). These operations yield:

$$(Y_{jt} - \bar{Y}_j - \bar{Y}_t + \bar{\bar{Y}}) = \beta_1 (X_{1jt} - \bar{X}_{1j} - \bar{X}_{1t} + \bar{\bar{X}}_1) + \ldots + \beta_K (X_{Kjt} - \bar{X}_{Kj} - \bar{X}_{Kt} + \bar{\bar{X}}_K). \quad (5)$$

If we let

$$W_{jt} = Y_{jt} - \bar{Y}_j - \bar{Y}_t + \bar{\bar{Y}} \quad (j = 1, \ldots J ; t = 1, \ldots, T)$$

and

$$Q_{ijt} = X_{ijt} - \bar{X}_{kj} - \bar{X}_{kt} + \bar{\bar{X}}_i \quad (i = 1, \ldots, K),$$

Equation (5) becomes

$$W_{jt} = \beta_1 Q_{1jt} + \beta_2 Q_{2jt} + \ldots + \beta_K Q_{Kjt}. \tag{6}$$

Equation (6) can then be estimated using least squares by forcing the mean to zero.

<u>A cross-sectionally heteroskedastic and time-wise autoregressive model</u>

In OLS we normally assume that (1) cross-sectional data are mutually independent and homoskedastic, and (2) time-series observations are homoskedastic and non-autoregressive. Given a time-series of cross-sectional observations, relaxing the homoskedastic assumption for cross-sectional observations and the non-autoregressive assumption for time-series observations would produce a pooled time-series and cross-sectional model. To estimate such a model we adjust the variance-covariance matrix of the distrubances to render the least squares method tenable.

Consider writing Equation (1) in matrix notation:

$$Y = X\beta + \varepsilon \tag{7}$$

where

$$Y = \begin{bmatrix} Y_{11} \\ Y_{12} \\ \cdot \\ \cdot \\ \cdot \\ Y_{jt} \\ \cdot \\ \cdot \\ \cdot \\ Y_{JT} \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{1,11} & X_{2,11} & \cdots & X_{K,11} \\ 1 & X_{1,12} & X_{2,12} & \cdots & X_{K,12} \\ \cdot \\ \cdot \\ \cdot \\ 1 & X_{1,jt} & X_{2,jt} & \cdots & X_{K,jt} \\ \cdot \\ \cdot \\ \cdot \\ 1 & X_{1,JT} & X_{2,JT} & \cdots & X_{KJT} \end{bmatrix}, \quad \beta = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_K \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{1T} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{JT} \end{bmatrix}$$

In its most general form, the variance-covariance matrix of the disturbances, Ω , can be represented as follows:

$$\Omega = \begin{bmatrix} E(\varepsilon_{11}^2) & E(\varepsilon_{11}\varepsilon_{12} & \cdots & E(\varepsilon_{11}\varepsilon_{1T})E(\varepsilon_{11}\varepsilon_{21}) & \cdots & E(\varepsilon_{11}\varepsilon_{JT}) \\ E(\varepsilon_{12}\varepsilon_{11}) & E(\varepsilon_{22}^2) & & & & E(\varepsilon_{12}\varepsilon_{JT}) \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ E(\varepsilon_{1T}\varepsilon_{11}) & E(\varepsilon_{1T}\varepsilon_{12}) & & & & E(\varepsilon_{1T}\varepsilon_{JT}) \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ E(\varepsilon_{JT}\varepsilon_{11}) & E(\varepsilon_{JT}\varepsilon_{12}) & & & & E(\varepsilon_{JT}^2) \end{bmatrix},$$

where E denotes the mathematical expectations operator.

Relaxing the homoskedasticity assumption of the linear model for cross-sectional observations and the non-autoregression assumption for time-series observations, we specify the following in Ω:

(a) $E(\varepsilon_{jt}^2) = \sigma_j^2$ (hyteroskedasticity),

(b) $E(\varepsilon_{jt}\varepsilon_{mt}) = 0$ for all $j \neq m$ (mutual independence), and

(c) $\varepsilon_{jt} = \rho_j \varepsilon_{j,t-1} + U_{jt}$ (autoregression),

where

ρ_j = the autoregressive parameter in the disturbances and is unique to sample unit j;

$\mu_{jt} \sim N(0, \sigma_u^2)$;

$\varepsilon_{jo} \sim N(0, \frac{\sigma_u^2}{1-\rho_j^2})$, and

$E(\varepsilon_{j,t-1} U_{mt}) = 0$ for all j and m.

With these specifications and assumptions, we can rewrite the Ω matrix as a [(J x T) x (J x T)] diagonal matrix with $\sigma_j^2 W_j$ as a typical element in the diagonal, j = i, ..., J. Each σ_j^2 stands for the variance of the disturbances pertaining to the j^{th} sample unit and each W_j is itself a T x T matrix as follows:

$$W_j = \begin{bmatrix} 1 & \rho_j & \rho_j^2 & \rho_j^{T-1} \\ \rho_j & 1 & \rho_j & \rho_j^{T-2} \\ . & . & 1 & . \\ . & . & . & . \\ . & . & . & . \\ \rho_j^{T-1} & \rho_j^{T-2} & \rho_j^{T-3} & 1 \end{bmatrix}$$

This model for the p/e ratio analysis, say the forces that determine each firm's p/e ratio, do change over time, but there are characteristics unique to each firm that cause these factors to shift differently for each firm. For example, firm X might be in an industry that is more sensitive to environmental changes than firm Y. Thus, in each time period the probability of random changes in firm X's p/e ratio is higher than that of firm Y. In other words, the probability distribution of p/e ratio for firm X is different from that of firm Y. The homoskedasticity assumption is thus relaxed. Furthermore, the environmental forces themselves shift from period to period, causing dynamic changes to all forces determining p/e ratios. Yet we assume that these dynamic forces have different impacts on each firm's p/e ratios. To this effect, we have specified ρ_j for each firm j to correct for the autoregression in the disturbance term. Note that this is the basic conceptual difference between the covariance model and this model because the former assumes that the dynamic forces affect all units equally, designated by c_t for all units at time t.

The computational procedures are described in Kmenta [9; 510-12]. Briefly, they are:

(i) Apply OLS to all J x T observations.

(ii) Use the resulting residuals to obtain consistent estimates of ρ_j's.

(iii) Perform the Orcutt transformation for all observations using the ρ_j's.

(iv) Run OLS using the transformed data.

(v) Use the resulting residuals to obtain consistent estimates of σ_j^2.

(vi) Perform weighted least squares using $\hat{\sigma}_j^2$ from step (v).

The resulting error term in the estimating equation can be shown to be asymtotically non-autoregressive and homoskedastic, thus rendering OLS appropriate. A full discussion of the statistical properties of the model can be found in Kmenta [9].

Choice of Variables

To use the models discussed, it is still necessary to specify the explanatory variables. Because this is an attempt to improve upon the methodology rather than an effort to discover new explanatory variables, we resorted to a theoretical formulation of the relationship discussed elsewhere and took a naive approach to specify the variables according this formulation. (Readers who are interested in knowing the dependent variables used in previous studies are referred to Keenan [8] and Cohen and Smyth [3].)

Lerner and Carleton [10] have formulated the following relationship:

$$p = \frac{d}{(k-g)}$$

where p = stock price,

k = the market discount rate,

d = per share dividends, and

g = expected earning growth.

If d = (1-r) e, where r equals the earnings retention rate and e equals earnings, we can rewrite the formulation as

$$p/e = \frac{(1-r)}{(k-g)} .$$

According to this formulation, the p/e ratio should be directly proportional to the payout rate and earnings growth, and inversely proportional to the discount rate.

Our final specification is as follows:

$$(p/e)_{jt} = \alpha + \beta_1 \text{ (Payout Rate)}_{jt} + \beta_2 \text{ (Share Price)}_{jt} + \beta_3 \text{ (EPS}_t - \text{EPS}_{t-1})_j + \beta_4 \text{ (Price}_t - \text{Price}_{t-1})_j . \quad (8)$$

As in previous studies, proxies for risk and growth are included. These proxies, variation in earnings per share (EPS) and share prices, are represented by the first difference of the two series.

All the data used in this analysis come from the Quarterly Compustat Tapes. The earnings-per-share variable includes extraordinary items. The share prices used are the month-end prices of the third month of each quarter. Data were obtained for 20 quarters extending from the first quarter of 1970 to the last quarter of 1974. The overall sample for the study consists of a randomly selected list of 150 firms with no missing data for any of the variables mentioned. Descriptive measures of these data are presented in Table 1. With the exception of the observations on price, all observations are skewed to the right. It is also obvious that the p/e variable has a flat-tailed distribution.

Table 1

DESCRIPTIVE STATISTICS OF RAW DATA

	N	Minimum	Mean	Maximum	Standard Deviation
p/e	3000	-1870.00	256.70	1.16×10^5	3389.70
Price	3000	.30	21.15	431.40	29.31
Difeps	3000	- 13.17	.0077	13.05	.831
Difprice	3000	- 132.60	-.366	66.20	6.85
Payout Rate	3000	- 26.25	3.455	1250.0	47.95

Results

The analysis consists of three stages. First, OLS was applied to the raw data in order to evaluate criticisms of previous studies that cross-sectional analysis resulted in coefficients that not only vary from period to period, but also are sample sensitive. Second, covariance regression was performed on different samples to determine if it would minimize sample sensitivity. Third, the heteroskedastic and autoregressive model was applied to the data for the same purpose.

Stage 1 yielded two sets of results. In one case, OLS was applied to five randomly selected subsamples for each of the five most recent quarters, yielding twenty-five sets of coefficients, presented in Table 2. These results tend to agree with existing criticisms about unstable parameters over time and in different samples. As Table 2 shows, the payout rate appears to be a significant explanatory variable consistently in all the regressions, but its coefficient falls short of being stable over time and across samples. The payout rate coefficient ranges

Table 2

ORDINARY LEAST SQUARES RESULTS (five different samples, five quarters)

	Quarter	Constant	Payout Ratio	Price	Difeps*	Difpr*	R^2
Sample 1 N = 75 firms	16	5.464	61.75				.847
		(2.042) δ	(3.066)				
	17	6.748	46.58	1.022			.854
		(2.015)	(2.389)	(.1136)			
	18	24.786	49.828				.245
		(11.280)	(10.252)				
	19	9.674	56.99			1.432	.407
		(4.049)	(8.507)			(.614)	
	20	--	50.710				.933
			(1.607)				
Sample 2 N = 75 firms	16		41.507				.999
			(.33)				
	17		48.165	3.169			.792
			(3.811)	(.357)			
	18		42.301	2.981			.854
			(2.263)	(.350)			
	19		65.592	2.458			.446
			(12.392)	(.503)			
	20		40.156				.355
			(6.385)				
Sample 3 N = 50 firms	16		41.486				.999
			(.052)				
	17		45.184				.841
			(2.861)				
	18		43.218	1.62		3.411	.890
			(2.39)	(.505)		(1.648)	
	19	15.665	36.199			1.905	.265
			(10.394)			(.823)	
	20		91.387				.454
			(8.243)				

Table 2 (continued)

	Quarter	Constant	Payout Ratio	Price	Difeps*	Difpr*	R^2
Sample 4 N = 50 firms	16	11.057 (2.758)	41.543 (5.103)				.579
	17	9.467 (4.714)	38.584 (6.639)				.413
	18	11.646 (4.757)	63.714 (5.769)	1.370 (.288)			.753
	19	26.157 (8.591)	53.250 (14.423)				.221
	20						
Sample 5 N = 50 firms	16		65.154 (2.509)	1.476 (.236)		-14.188 (5.324)	.941
	17		47.028 (2.544)	1.654 (.242)			.897
	18	32.733 (16.594)	48.587 (15.046)				.178
	19		62.481 (10.980)	.928 (.272)			.470
	20	13.496 (3.586)	50.642 (1.882)				.939

*Difeps and Difpr are the first difference in earnings per share and the first difference in price, respectively.

$^{\delta}$Values in parenthesis are standard errors of the coefficients.

---

from 36.199 (sample 3, quarter 19) to 91.387 (sample 3, quarter 20). The other coefficients showed even more variation over time and across samples. The stock price coefficient appeared to be almost consistently significant on in samples 2 and 5. On the whole, however, no définitive statement can be made about the significance of the stock price variable, Difeps, and Difpr. The R^2 values also had a wide range, from as low as 0.178 to a high of .999. That the explanatory power of the explanatory variables do vary over time can be inferred from the shifts in R^2 over time within each sample.

Before we left the first stage, we attempted to further focus on the instability of coefficients over time. Regressions were run for each of the 20 quarters for the entire sample of 150 firms. The results are presented in Table 3. Although the share price variable does not seem to vary too much whenever it is significant (0.81 to 1.48), it is not consistently significant over time. As for the payout rate, the instability of the coefficients is evidenced by its range from 30.86 to 191.25, as is illustrated in the histogram in Figure 1. Also worth noting are the R^2 values over time which range from .42 to .99. The variation of R^2 values clearly shows that the environmental influences on p/e determination do change from period to period. If we ignore these influences by employing OLS alone, it is guaranteed that no consistent models can result from the analysis.

In sum, these regressions indicate that payout rate is an important explanatory variable for p/e ratios. Because of short-run fluctuations in exogeneous forces, other variables specified here tend to assume significance, but not consistently over time. If we wish to address the question of identifying determinants of p/e ratios in the long-run, the above results do not produce a single set of answers.

The covariance regression was the second stage of the analysis. Covariance regression was run on five randomly selected subsamples from the 150 sample firms. The results, presented in Table 4, show R^2 values

Table 3

REGRESSION RESULTS FOR TWENTY QUARTERS

Quarter	Constant	Payout Ratio	Price	Difeps	Difpr	R^2
1		91.387 (8.243)				.454
2		85.575 (5.1198)				.655
3		58.653 (.212)				.998
4		54.384 (.428)				.991
5		189.06 (9.153)	1.472 (.366)	26.84 (6.86)		.769
6		56.84 (.395)				.993
7	59.573 (9.252)	57.434 (.161)	.949 (.245)			.999
8	48.427 (8.434)	56.397 (.182)				.998
9		191.25 (14.619)	1.478 (.231)			.571
10	43.046	57.74 (.106)	.814 (.167)			.999
11	29.053 (6.218)	55.52 (.127)				.999
12	33.325 (4.011)	30.855 (0.043)				.999
13		118.98 (6.969)	1.187 (.269)		-5.258 (1.114)	.699
14	23.013 (5.502)	49.97 (.101)				.999
15	13.407 (3.218)	40.5 (.0434)	.872 (.156)			.999
16		41.496 (.096)				.999

Table 3 (continued)

Quarter	Constant	Payout Ratio	Price	Difeps	Difpr	R^2
17	7.29	48.527	1.203			.769
	1.854	(2.344)	(.123)			
18	19.053	45.190				.350
	(5.849)	(5.062)				
19	14.65	69.463	1.181			.418
	(3.717)	(7.523)	(.190)			
20	21.413	43.965				.492
	(6.819)	(3.696)				

Sample size = 150 Firms

```
MIDPOINT     COUNT FOR 1.V1  (EACH X= 1)

 30.860          1 +X
 39.302          2 +XX
 47.743          4 +XXXX
 56.185          7 +XXXXXXX
 64.626          0 +
 73.068          1 +X
 81.509          1 +X
 89.951          1 +X
 98.393          0 +
 106.83          0 +
 115.28          1 +X
 123.72          0 +
 132.16          0 +
 140.60          0 +
 149.04          0 +
 157.48          0 +
 165.93          0 +
 174.37          0 +
 182.81          0 +
 191.25          2 +XX

 TOTAL          20  (INTERVAL WIDTH= 8.4416)
```

Fig. 1. Histogram showing frequencies of coefficient of p/e rate for different quarters.

Table 4

COVARIANCE REGRESSION RESULTS

	Payout Ratio	R^2	SE
Sample 1 N = 18 firms	97.967 (2.481)	.813	167.56
Sample 2 N = 13 firms	98.898 (2.910)	.819	167.06
Sample 3 N = 43 firms	96.725 (1.70)	.791	180.22
Sample 4 N = 75 firms	97.716 (1.362)	.782	192.51
Sample 5 N = 75 firms	97.286 (1.323)	.788	185.67

staying relatively stable for the five samples. It turned out, however, that only the payout rate was significant for all five samples. We shall discuss this point later. We can say, however, that since the coefficients do not vary from sample to sample the results support the claim that pooling time-series and cross-sectional observations is more appropriate in estimating equilibrium conditions.

The third stage used the heteroskedastic-and-autoregressive model. The model was run on four subsamples from the same 150 sample firms. Results are presented in Table 5 which shows that at least two variables, price and payout rate, are consistently significant for different samples. What is disappointing, however, is that the two pooled

Table 5

HETEROSKEDASTIC-AND-AUTOREGRESSIVE MODEL RESULTS

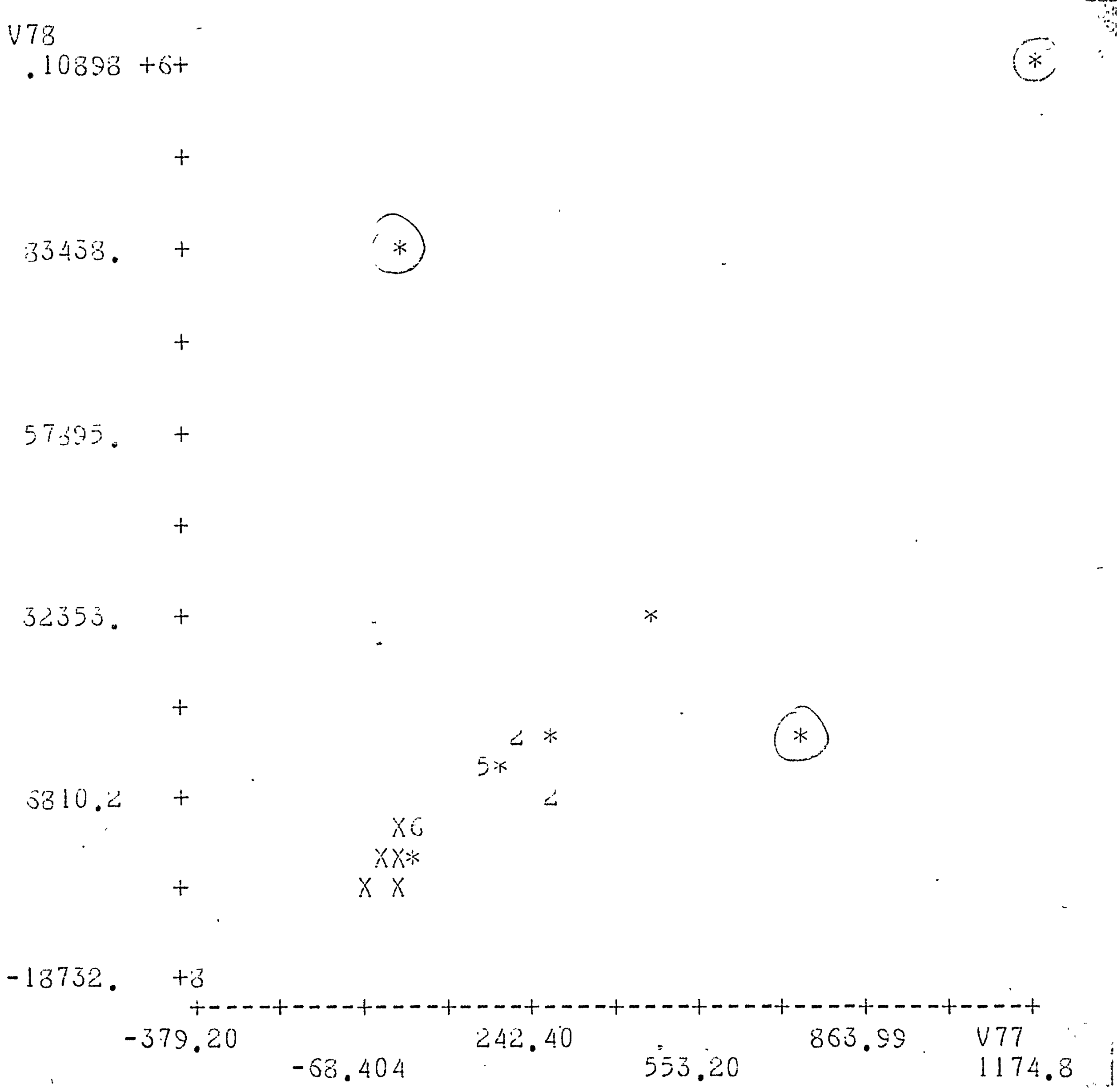
	Price*	Difpr*	Payout*	R^2
Sample 1 N = 34 firms	.997 (.037)	.156 (.045)	51.379 (2.481)	.908
Sample 2 N = 31 firms	.802 (.048)		59.652 (3.133)	.772
Sample 3 N = 29 firms	1.023 (.035)	.271 (.047)	50.078 (2.284)	.938
Sample 4 N = 31 firms	.881 (.044)		57.431 (3.404)	.736

models yielded quite different results. It was this discrepancy that led us to suspect that the outliers in the data are so extreme as to distort the models. We shall expose the problem in the next section and will examine it in greater detail in a future paper.

Problem of Extreme Outliers

Because the results from both of the pooled models are not satisfactory, this section will explore the potential problem caused by outliers. It will attempt to offer some explantion for the two pooled models resulting in different sets of estimates.

Bypassing the question of omitted variables, which will be handled in the Discussion section, we will first examine the structure of the data. Figure 2 plots the transformed observations of p/e ratios and payout rates used in the covariance regression. It is obvious that the results might have been biased by the three outliers circled in the



* = one observation.

X = more than ten points.

Fig. 2. Scatter plot of transformed p/e and payout rate.

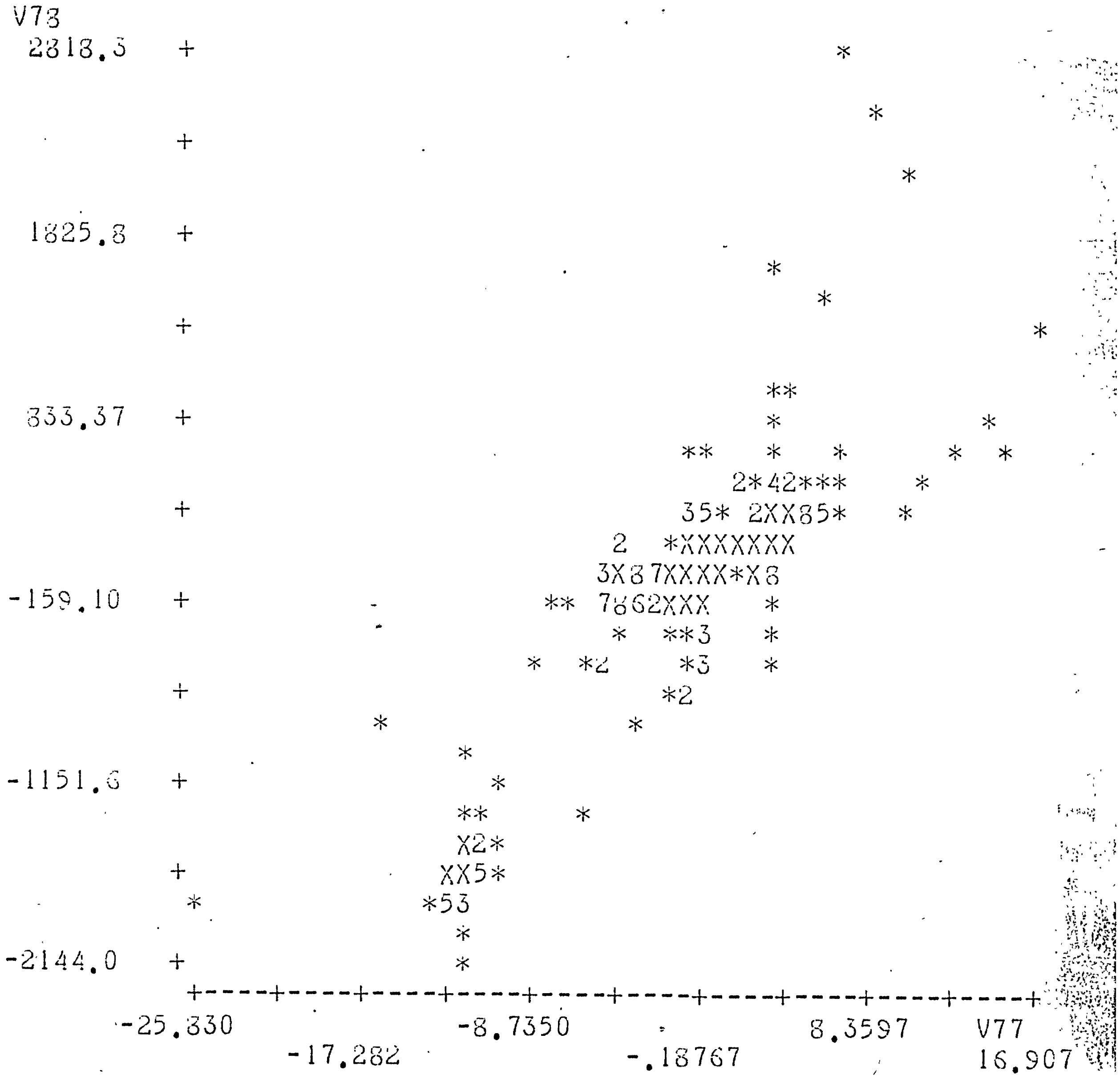
plots. Accordingly, in Figure 3 we replotted all observations except the three apparent outliers, with much improvement. A study of the second plot reveals that the data actually have two groupings. Because we do not know the exact counts of the "x" in the graph beyond that its greater than 10, the plot does not indicate whether the bi-modal data structure might be a problem.

Further evidence of outlier bias is shown in Figure 4, where the residuals from a covariance regression using all sample data are plotted against the transformed payout rates. Again, three outliers are prominent in the plot. By discarding these three points we obtained a much better plot as presented in Figure 5. Finally we arrayed the transformed payout rate data in Figure 6 and determined that, indeed, the explanatory variable has too many outliers.

For the heteroskedastic-and-nonautoregressive model, we ran a regression using all observations then plotted the residuals against adjusted share prices and payout rates. These plots are presented in Figures 7 and 8. Again, the effects of the outliers are conspicuous, as evidenced by only one "x" in each graph, which represents more than 99 percent of the observations. Thus, the impact of the central cloud of data, about which we would like to learn more, has been mitigated by the extreme outliers.

Discussion

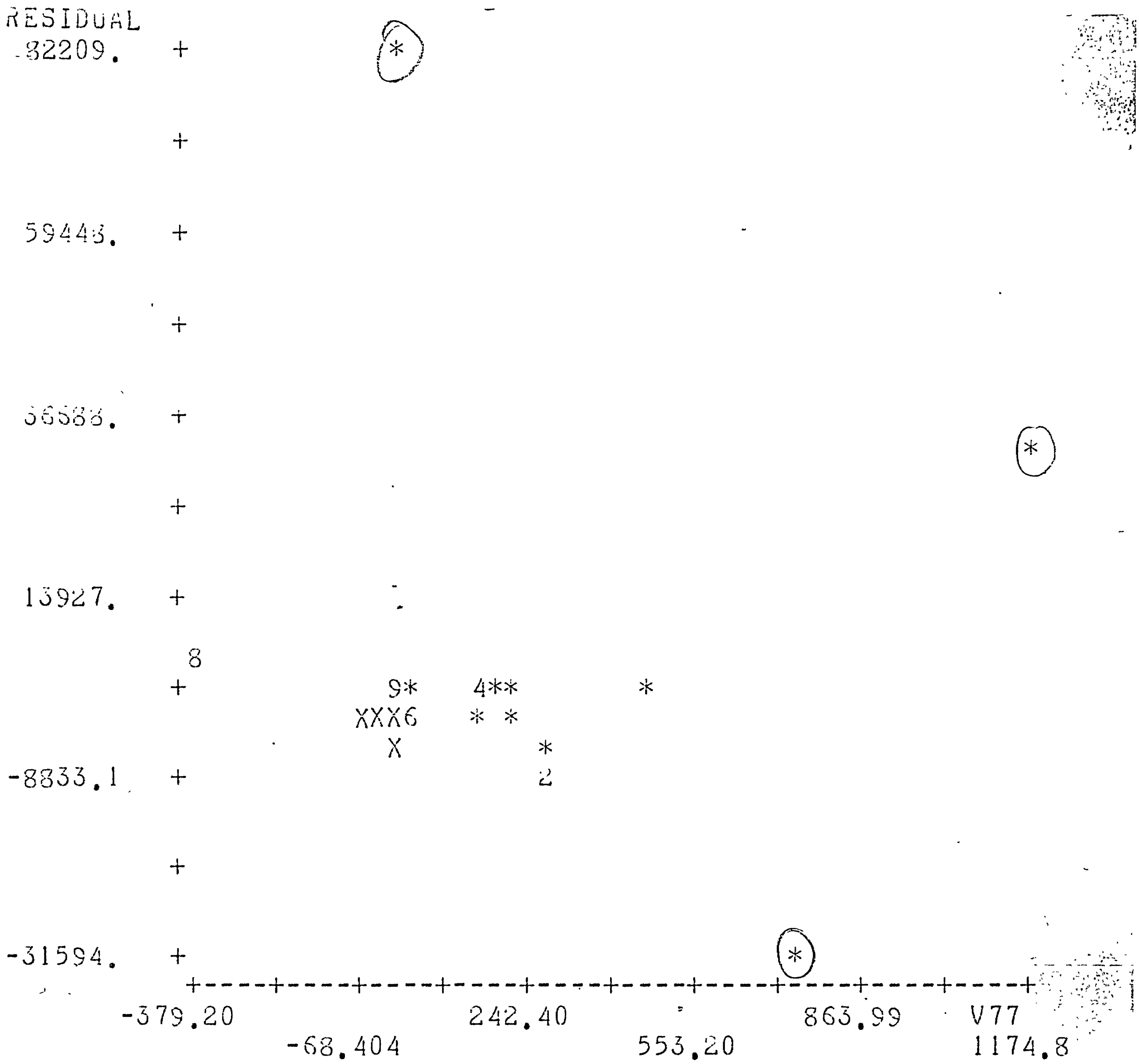
Study of these data increase our understanding of why previous p/e ratio studies failed to supply a consistent set of determinants. This



* = one observation.

X = more than ten points.

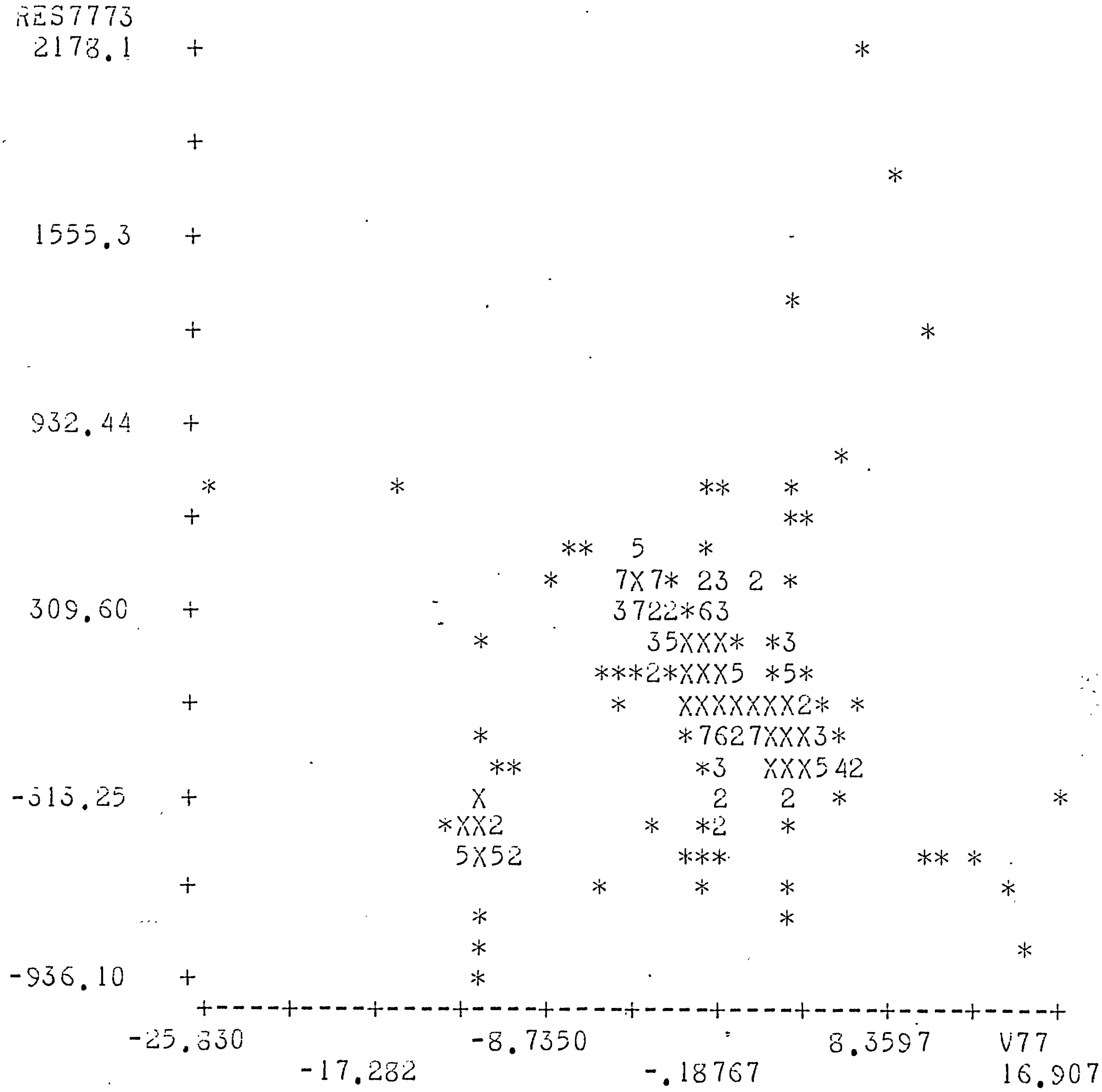
Fig. 3. Scatter plot of transformed p/e and payout rate net of three outliers.



* = one observation.

X = more than ten points.

Fig. 4. Scatter plot of residuals from covariance regression and payout rate.



* = one observation.

X = more than ten points.

Fig. 5. Scatter plot of residuals from covariance regression and payout rate net of three outliers.

MIDPOINT		(EACH X= 74)
-379.20	8	+X
-347.49	0	+
-315.78	0	+
-284.06	0	+
-252.35	0	+
-220.63	0	+
-188.92	0	+
-157.20	0	+
-125.49	0	+
-93.776	0	+
-62.061	19	+X
-30.347	21	+X
1.3672	2923	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
33.081	0	+
64.796	0	+
96.510	0	+
128.22	1	+X
159.94	4	+X
191.65	1	+X
223.37	2	+X
255.08	0	+
286.80	3	+X
318.51	0	+
350.22	0	+
381.94	0	+
413.65	0	+
445.37	0	+
477.08	1	+X
508.80	0	+
540.51	0	+
572.22	0	+
603.94	0	+
635.65	0	+
667.37	0	+
699.08	0	+
730.79	0	+
762.51	1	+X
794.22	0	+
825.94	0	+
857.65	0	+
889.37	0	+
921.08	0	+
952.79	0	+
984.51	0	+
1016.2	0	+
1047.9	0	+
1079.7	0	+
1111.4	0	+
1143.1	0	+
1174.8	1	+X

Fig. 6. Frequency distribution of the transformed payout rate variable.

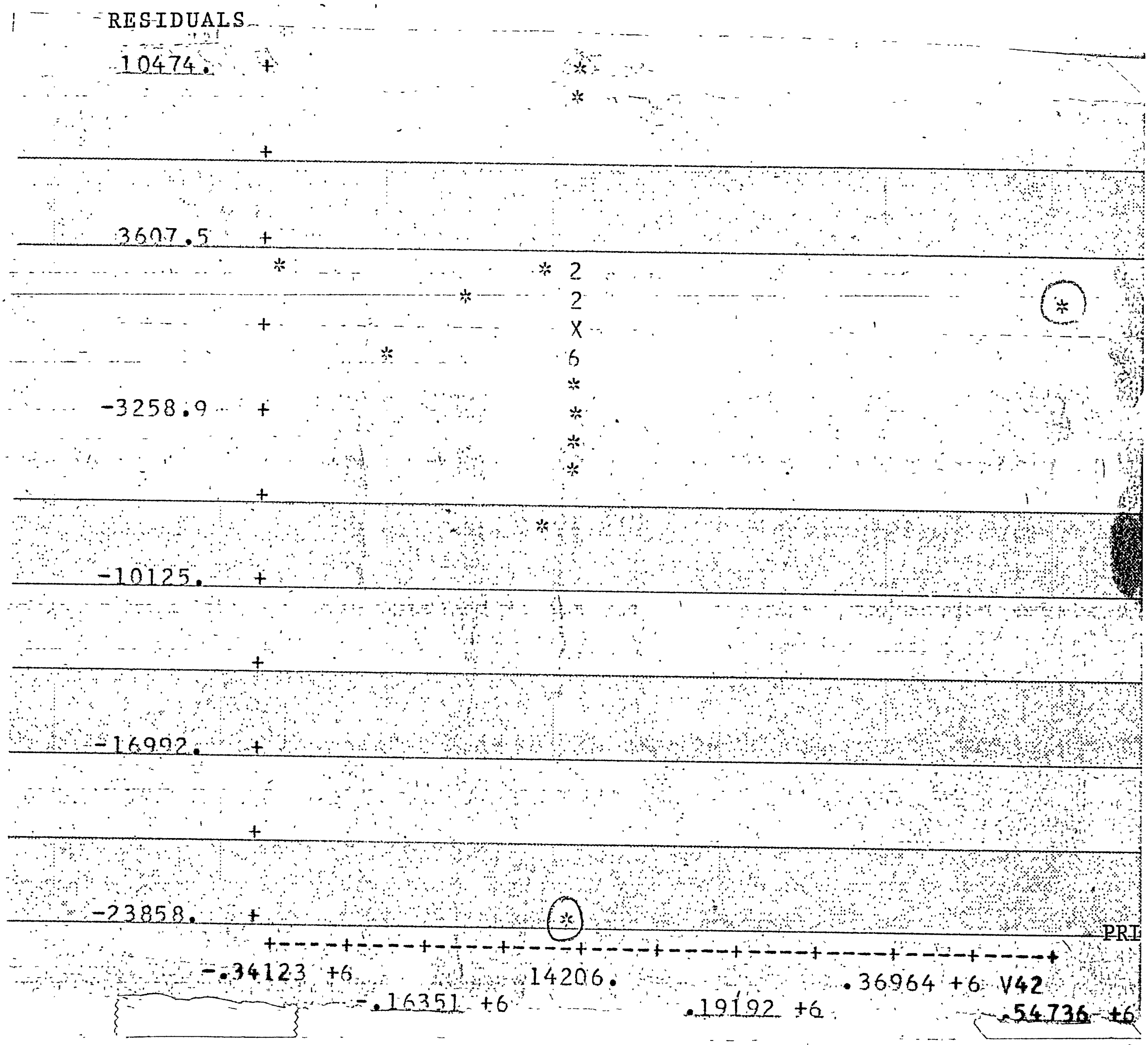


Fig. 7. Scatter plot of residuals from the second pooled model and adjusted price variable.

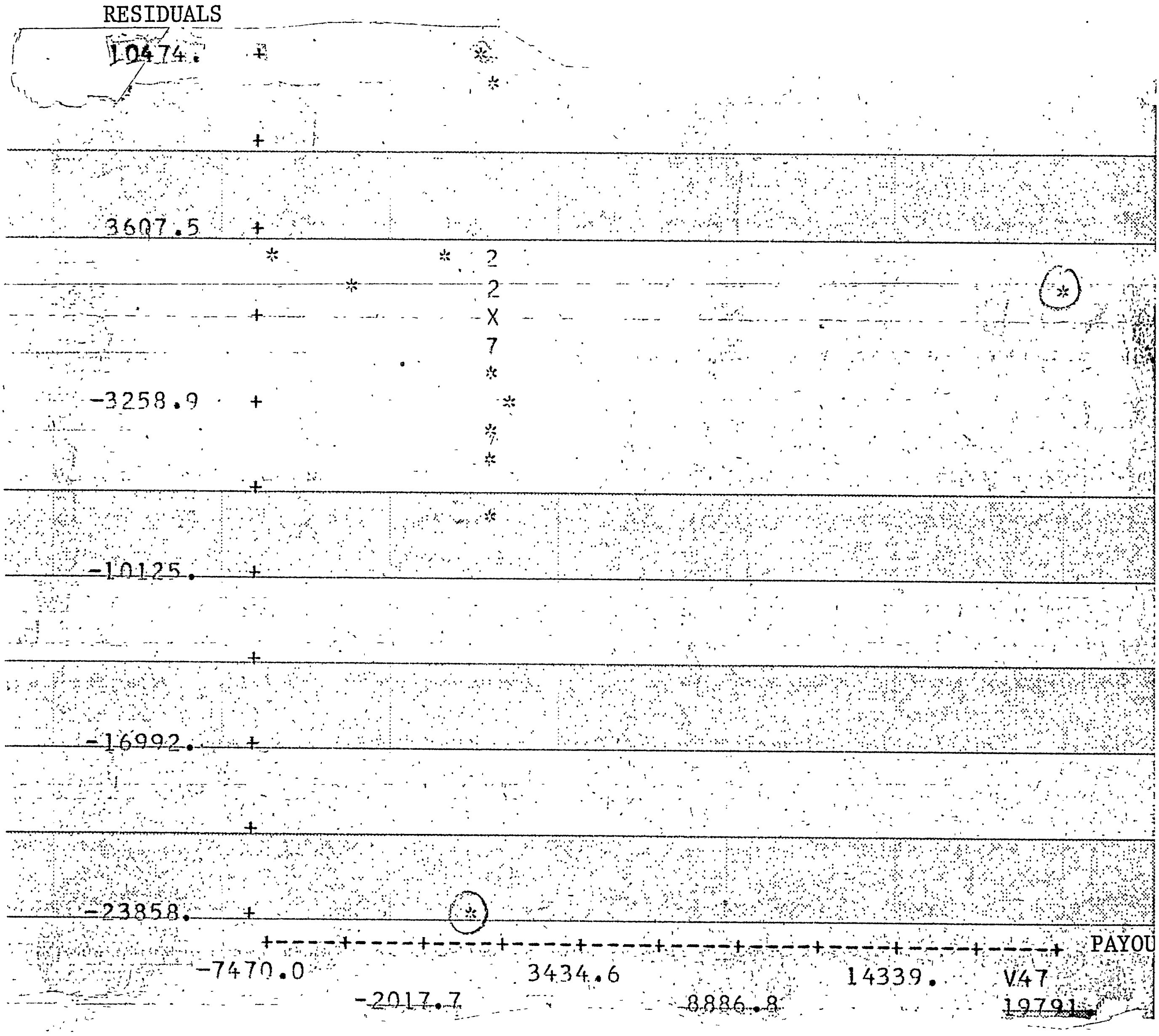


Fig. 8. Scatter plot of residuals from the second pooled model and adjusted payout rate.

study has illustrated that the use of pooling time-series of cross-sectional information can add to the estimation of relationships between variables when one suspects that dynamic, unspecified, exogeneous influences exist. Even with random sampling, the data have too many outliers which are likely to bias the estimates. In general, this study has provided supporting evidence for the contention that not only are parameter values of variables unstable over time, but also the number variables having significant explantory power actually shift from period to period.

Each pooled regression model yields consistent results for different samples, which supports the hypothesis of this study. One can consider cross-sectional analysis as accounting for quasi-equilibrium patterns of relationship between variables, whereas time series data reflect responses to changes in the accustomed levels of the independent variables. Thus, according to our results, in the process of moving toward equilibrium conditions, p/e ratios will continue to be determined, in aggregate, as a function of the payout rate.

Unfortunately, we obtained two different sets of results for the two pooled models. The covariance model indicates that, when the payout rate increases by one percentage point, the corresponding p/e ratio will adjust itself and eventually increase by one time. On the other hand, the heteroskedastic-and-autoregressive model suggest that p/e ratios are a function of both price and payout rate. If price increases by one dollar, *ceteris paribus*, the corresponding p/e multiple will increase by one time. Similarly, when the payout rate increases by 1 percent, the corresponding p/e ratio will increase by one-half of 1 percent.

Whether this finding is consistent with the "true" formula for determining p/e ratios is uncertain. Given the potential flaws in the models, this analysis has provided additional evidence to support the claim that dividends are relevant in the valuation of capital assets.

As an extension of this present effort, a possible area for further investigation would be to extract the b_j's from the covariance model to examine their relationship with firm characteristics. Similarly, the c_t's may be capable of illuminating the exogeneous effects in different time periods. This methodology can be applied to many other areas besides p/e ratios.

Footnotes

1. A cursory review of the literature will demonstrate that this is the case. Especially in the area of efficient market research, OLS is the fundamental statistical model. For example, see [1] and [5].

2. The importance of exploratory data analysis and robust techniques have been summarized by John Tukey [12]. Interested readers can get a good exposure to robust estimation from Huber [6].

3. See Bing [2].

References

1. Ball, R., and Brown, P. "An Empirical Evaluation of Accounting Income Numbers." Journal of Accounting Research, Autumn 1968,, pp. 159-78.

2. Bing, Ralph A. "Survey of Practitioners' Stock Evaluation Methods." Financial Analysts Journal 27, No. 3 (May-June 1971): 55-60.

3. Cohen, Stanley, and Smyth, David J. "Some Determinants of Price/Earnings Ratios of Industrial Common Stocks." Quarterly Review of Economics and Business 13, No. 4 (Winter 1973): 49-60.

4. Fama, E. "The Behavior of Stock Market Price." Journal of Business, January 1965A: 34-105.

5. Fama, E., Fisher, L., Jensen, M., and Roll, R. "The Adjustment of Stock Prices to New Information." International Economic Review, February 1969: 1-21.

6. Huber, P. J. "The 1972 Wald Lecture, Robust Statistics: A Review." The Annuals of Mathematical Statistics 43, No. 4 (1972): 1041-67.

7. Johnson, K. H., and Lyon, H. L. "Experimental Evidence on Combining Cross-Section and Time Series Information." Review of Economics and Statistics LV, No. 4 (November 1973): 465-74.

8. Keenan, Michael. "Models of Equity Valuation: The Great SERN Bubble." Journal of Finance 25 (May 1970): 243-65.

9. Kmenta, J. "Elements of Econometrics." New York: Macmillan and Company, 1971.

10. Lerner, E. M., and Carleton, W. T. "Financing Decisions of the Firm." Journal of Finance 21 (May 1966): 202-14.

11. Schipper, Lewis. Consumer Discretionary Behavior. Amsterdam: North-Holland Publishing Company, 1964.

12. Tukey, John W. "Data Analysis, Computation and Mathematics." Quarterly of Applied Mathematics, April 1972: 51-65.