LINEAR REGRESSION DIAGNOSTICS*

Roy E. Welsch
and
Edwin Kuh

Massachusetts Institute of Technology
and
NBER Computer Research Center

Working Paper No. 173

PRELIMINARY DRAFT

Revised 25 March 1977

*We are indebted to the National Science Foundation for supporting this research under Grant SOC76-14311 to the NBER Computer Research Center

ABSTRACT

This paper attempts to provide the user of linear multiple regression with a battery of diagnostic tools to determine which, if any, data points have high leverage or influence on the estimation process and how these possibly discrepant data points differ from the patterns set by the majority of the data. The point of view taken is that when diagnostics indicate the presence of anomolous data, the choice is open as to whether these data are in fact unusual and helpful, or possibly harmful and thus in need of modifications or deletion.

The methodology developed depends on differences, derivatives, and decompositions of basic regression statistics. There is also a discussion of how these techniques can be used with robust and ridge estimators. An example is given showing the use of diagnostic methods in the estimation of a cross-country savings rate model.

ACKNOWLEDGEMENTS

The authors would like to acknowledge helpful conversations with David Hoaglin, Frank Hampel, Richard Hill, David Andrews, Jim Frane, Colin Mallows, Doug Martin, and Fred Schweppe.

TABLE OF CONTENTS

		Page
1.	INTRODUCTION	1
1.1	General Goals	1
1.2	Regression Diagnostics and Model Input Perturbations	3
1.3	Modeling Research Aims and Diagnostics	6
1.3.1	Leverage and Disparate Data	6
1.3.2	Collinearity	7
1.3.3	Regression Parameter Variability in Time	8
1.4	Notation	9
2.	LEVERAGE POINTS AND DISPARATE DATA	10
2.1	Introduction	10
2.2	Residual Diagnostics	14
2.3	The Hat Matrix	18
2.4	Row Deletion Diagnostics	23
2.5	Regression Statistics	26
2.6	Influence and Variance Decomposition	28
2.7	More Than One Row at a Time	31
2.8	Interface with Robust and Ridge Regression	33
2.9	An Example: An Inter-Country Life Cycle Savings Function	36
2.9.1	Residuals	38
2.9.2	Leverage and Diagonal Hat Matrix Entries	38
2.9.3	Coefficient Perturbation	39
2.9.4	Variation in Coefficient Standard Errors	39
2.9.5	Change in Fit	40
2.9.6	A Provisional Summary	41
2.9.7	One Further Step	42
2.10	Final Comments	43
REFERENCES		44
Appendix 1.	BASIC DIFFERENCE FORMULAS	A1.1
Appendix 2.	DIFFERENTIATION FORMULAS	A2.1
Appendix 3.	THEOREMS ON THE HAT MATRIX	A3.1
Appendix 4.	EXHIBITS FOR SECTION 2.9	A4.1

1. INTRODUCTION

1.1 General Goals

Economists and other model builders have responded willingly to major opportunities that have appeared in the past two decades - a rapidly growing demand for policy guidance and forecasts from government and business, and the purely intellectual goal of advancing the state of knowledge through model development. The fundamental enabling condition has been the ability to produce more intricate models at decreasing unit cost because of advances in computer technology. A large econometric model twenty years ago had twenty equations: today a large model has a thousand equations. It is not only larger models, but also larger data sets and more sophisticated functional forms and estimators that have burgeoned.

The transition from sliderule and desk calculator to the large scale digital computer has happened with startling speed. The benefits have, in our opinion, been notable and at times exciting; we know a great deal more about the economy and can provide more intelligent guidance as a direct result of increased computational power. At the same time, there are hidden costs of current approaches to quantitative economic research via computer which ought to be recognized.

One major cost is that, today, the researcher is a great deal further away from data than he was, perforce, in the heyday of the desk calculator. If there are a great many equations to estimate or thousands of observations for a few equations, there is a natural tendency to use the computer for what it does well: process data. A tape arrives and after a frustrating day or two is accessible by a computer program (often a regression package, plain or

fancy). Then estimation and hypothesis testing get underway until some satisfactory conclusion is obtained. It is not misguided nostalgia to point out that it was more likely, with the more labor intensive technology of the past, for the researcher to uncover peculiarities in the data. Nor do we counsel a return to the golden past. What concerns us is that the "something" which has been lost in modern practice is valuable and is not recoverable from standard regression statistics. Our first major objective is to suggest procedures that exploit computer brawn in new ways that will permit us to get closer to the character of the data and its relation to hypothesized and estimated models.

There is the related issue of reliability. Our ability to crunch large quantities of numbers at low cost makes it feasible to iterate many times with a given body of data until the estimated model meets widely accepted performance criteria in terms of statistical measures such as t statistics, Durbin-Watson statistics and multiple correlations, along with theoretically approved coefficient signs and magnitudes. The iterative process is not what the statistical theory employed was originally all about, so that it behooves us to consider alternative ways of assessing reliability, which is a second major objective of this paper.

Another aspect of reliability is associated with questions of distance from the data that were mentioned at the outset. Specifically, the closer one is to the data, the more likely it is that oddities in the data will be uncovered or failure of the model and data to conform with each other will be discernible, so that reliability can be increased when the researcher has more intimate contact with the data. At the same

time, this posses a dilemma, since the researcher may then be excessively prone to devise theories from data. This temptation, often referred to as data mining, should be restrained. One sort of insurance against data mining is to be a strict Baysian and thus be guided by sensible rules for combining prior and posterior information. Alternatively the model should be tested - repeatedly if possible - on bodies of data unavailable at the time. Being a strict Baysian is not always practical nor is it deemed to be universally desirable. As a general rule then, the most practical safeguard lies with replication using previously unavailable data.

1.2 Regression Diagnostics and Model Input Perturbations

This paper presents a different approach to the analysis of linear regression. While we will sometimes use classical procedures, the principal novelty is greater emphasis on new diagnostic techniques. These procedures sometimes lack rigorous theoretical support, but possess a decided advantage in that they will serve as yet unmet needs of applied research. A significant aspect of our approach is the development of a comprehensive set of diagnostics.

An important underlying concept is that of perturbing regression model inputs and examining the model output response. We view model inputs broadly to include data, parameters (to be estimated), error models and estimation assumptions, functional form and a data ordering in time or space or over other characteristics. Outputs include fitted values of the dependent variable, estimated parameter values, residuals and functions of these (R^2, standard errors, autocorrelations, etc.).
We plan to develop various types of input perturbations that will reveal where model outputs are unusually sensitive. Perturbations can take the

form of differentiation or differencing, deletion (of data), or a change in estimation or error model assumptions.

The first approach to perturbation is "differentiation" (in a broad sense) of output processes with respect to input processes, in order to find a rate of change. This will provide a first order measure of how output is influenced by input; differences would be substituted for derivatives in discrete cases. If the rate of change is large, it can be a sign of potential trouble. Generally, one would like to have small input perturbations lead to small output deformations. We would also use this idea to see how big a perturbation can be before everything breaks down. Of course, a "good" model is generally responsive to anticipated changes in input.

For example, one could "differentiate" the model with respect to its parameters to ascertain output sensitivity to small changes in the parameters. (We could, for example, evaluate this parameter sensitivity function at the estimated parameter values.) This might indicate some of the more critical parameters in the model that deserve further analysis.

A second procedure is to perturb the input data by deleting or altering one data point and observe changes in the outputs. More generally we can remove random groups of data points or, for time series, sequences of data points. This is one way to search for parameter instability over time. By deleting individual data points or collections of points one can observe whether or not subsets of the data exert unusual influence on the outputs. In particular, it is possible to establish if a minority of the data behave differently from the majority of the data. The concept

of discrepant behavior by a minority of the data is basic to the diagnostic view elaborated in this paper.

The third approach will be to examine output sensitivity to changes in the error model. Instead of using least squares, estimators such as least absolute residuals would be applied which impute less influence to large residuals. A more promising alternative for diagnostic purposes is the Huber type error model [1]. Varying a parameter in the Huber model provides a way to examine sensitivity to charges in the error assumptions. This area is related to recent research in robust statistics[17].

Another aspect of changed error assumptions is specific to time series. Practicing econometricians are well aware that parameter estimates change when the sample period is altered. While this might only reflect expected sampling fluctuations, the possibility exists that the population parameters are truly variable and should be modeled as a random process. It is also possible that the population parameters are stable but mispecification causes sample estimates to behave as if they were a random process. In either case explicit estimation methods for random parameters based on the Kalman filter might reveal parameter instability of interest from a diagnostic point of view.

While classical statistical methods in most social science contexts treat the sample as a given and then derive tests about model adequacy, we take the more eclectric position that diagnostics might reveal weaknesses in the data, the model or both. Several diagnostic procedures, for example, are designed to reveal unusual rows or outliers in the data matrix which by assumption has no formal distribution properties. If a suspect data row has been located, the investigator faces several choices. One common practice is

to introduce a dummy variable, especially when subsequent examination reveals that an "unusual" situation could have generated that data row. Alternatively the model may be respecified in a more complex way. Of course the suspicious row might simply be deleted or modified if found to be in error. In summary, the diagnostic approach leaves open the question of whether the model, the data or both should be modified. In some instances described later on, one might discover a discrepant row and decide to retain it, while at the same time having acquired a more complete understanding of the statistical estimates relative to the data.

1.3 Modeling Research Aims and Diagnostics

We reiterate here several principal objectives that diagnostics can serve, from the modeler's perspective, in obtaining a clearer understanding of regression beyond those obtainable from standard procedures. Some of these are of recent origin or are relatively neglected and ought to be more heavily emphasized. The three main modeling goals are detection of disparate data segments, collinearity, and temporally unstable regression parameters. It will become clear as this paper proceeds that overlaps exist among detection procedures.

1.3.1 Leverage and Disparate Data

The first goal is the detection of data points that have disproportionate weight, either because error distributions are poorly behaved or because the explanatory variables have (multivariate) outliers. In either case regression statistics, coefficients in particular, may be heavily dependent on subsets of the data. (This draft is principally concerned with these

aspects of diagnosis; the other topics are of equal importance. At this stage of our research we are coming to a better understanding of the scope of regression diagnostics and we shall rely heavily on the work of others in describing these other methods.)

1.3.2 Collinearity

While exact linear dependencies are rare among explanatory variables apart from incorrect problem formulation, the occurance of near dependencies arises (all too) frequently in practice. While some collinearity can be moderated by appropriate rescaling, in many instances ill-conditioning remains. There are two separate issues, diagnosis and treatment. Since our main purpose is diagnosis, we are not presently concerned with what to do about it, except to note that the more collinear the data, the more prior information needs to be incorporated.

Collinearity diagnosis is experimental too, but the most satisfactory treatment we know of has been proposed by David Belsley [2], who builds on earlier work of Silvey [3]. By exploiting a technique of numerical analysts called the singular value decomposition, it is possible to obtain an index of ill-conditioning and relate this to a decomposition of the estimated coefficient variances. This relation enables the investigator to locate which columns of the explanatory variable matrix, associated with the index of collinearity, contribute strongly to each coefficient variance. By thus joining Silvey's decomposition of the covariance matrix to numerical measures of ill-conditioning, economists now have an experimental diagnostic tool that enables an assessment of which columns of the data matrix are prime sources of degradation in estimated coefficient variances.

1.3.3 Regression Parameter Variability in Time

A third major goal is the detection of systematic parameter variation in time. Many statistical models assume that there exist constant but unobservable parameters to be estimated. In practice, econometricians often find this assumption invalid. Suspicions that there are more than one set of population parameters can be aroused for a large number of reasons: the occurance of an external shock that might be expected to modify behavior significantly (a war, hyperinflation, price-wage controls, etc.) is one possibility. Another is that a poorly specified relation might exclude important variables which change abruptly. There is always the possibility that aggregation weights [4] may change over time and thereby introduce variability in macro parameters even when micro parameters are stable. An argument has been made by Lucas [23] that anticipated changes in government policy will cause modifications in underlying behavior. Finally the parameters may follow a random process and thus be inherently variable. When discrete changes in parameters are suspected, and the sub-divisions of data where this occurs is identifiable from outside information, the analysis of covariance in the form discussed in Gregory Chow [5] or Franklin Fisher [6] is an appropriate diagnostic that has been frequently applied. When the break point of points have to be estimated, maximum likelihood estimators proposed by Quandt and Goldfeld [7][8] are available.

An alternative diagnostic procedure has recently been suggested by Brown, Durbin and Evans [9]. They have designed two test statistics with a time series orientation. From a regression formed by cumulatively adding new observations to an initial subset of the data, one-step ahead predictions are generated. Both the associated cumulated recursive residuals

and their sums of squares have well-behaved distributions on the null hypothesis of parameter constancy.

1.4 Notation

We use the following notation:

Population Regression $Y = X\beta + \varepsilon$	Estimated Regression $Y = X\hat{\beta} + r$
Y : nx1 column vector for dependent variable	same
X : nxp matrix of explanatory variables	same
β : px1 column vector of regression coefficients	$\hat{\beta}$: estimate of β
ε : nx1 column error vector	r : residual vector
Additional notation	
x_i : i^{th} row of X matrix	same
σ^2 : error variance	s^2 estimated error variance
	$\hat{\beta}_{(i)}$ β estimated with i^{th} row of data matrix and Y vector deleted.

Other notation is either obvious or will be introduced in a specific context not so obviously tied to the generic regression model.

2. LEVERAGE POINTS AND DISPARATE DATA

2.1 Introduction

At this stage in the development of diagnostic regression procedures, we turn to analysis of the structure of the X matrix through perturbation of its rows. In the usual case, the X's are assumed to be a matrix of fixed numbers and the matrix to have full column rank. Otherwise, statistical theory suggests we ought to have little interest in the X matrix, except when experimental design considerations enter. In actual practice, researchers pay a great deal of attention to explanatory variables, especially in initial investigatory stages. Even when data are experimentally generated, peculiarities in the data can impact subsequent analysis, but when data are non-experimental, the possibilities for unusual data to influence estimation is typically greater.

To be more precise, one is often concerned that subsets of the data, i.e., one or more rows of the X matrix and associated Y's might have a disproportionate influence on the estimated parameters or predictions. If, for example, the task at hand is estimating the mean and standard deviation of a univariate distribution, exploration of the data will often reveal outliers, skewness or multimodal distributions. Any one of these might cast suspicion on the data or the appropriateness of the mean and standard deviation as measures of location and variability. The original model may also be questioned and transformations of the original data consistent with an alternative model may be suggested, for instance. In the more complicated multiple regression context, it is common practice to look at the univariate distribution of each column of X as well as Y, to see if any oddities (outliers or gaps) strike the eye. Scatter

diagrams are also examined. While there are clear benefits from sorting out peculiar observations in this way, diagnostics of this type cannot detect multivariate discrepant observations. That weakness is what we hope to remedy.

The benefits from isolating sub-sets of the data that might disproportionately impact the estimated parameters are clear, but the sources of discrepancy are diverse. First, there is the inevitable occurance of improperly recorded data, either at the source or in transcription to computer readable form. Second, observational errors are often inherent in the data. While more appropriate estimation procedures than least squares ought to be used, the diagnostics we propose below may reveal the unsuspected existance or severity of observational errors. Third, outlying data points may contain valuable information that will improve estimation efficiency. We all seek the "crucial experiment", which may provide indispensible information and its counterpart can be incorporated in non-experimental data. Even in this situation, however, it is constructive to isolate extreme points that indicate how much the parameter estimates lean on these desirable data. Fourth, patterns may emerge from the data that lead to a reconsideration and alteration of the initial model in lieu of suppressing or modifying the anomolous data.

Before describing multivariate diagnostics, a brief two dimensional graphic preview will indicate what sort of interesting situations might be subject to detection. We begin by an examination of Figure 1, which portrays the ideal null case of uniformly distributed and, to avoid statistical connotations, what might be called evenly distributed X. If the variance of X is small, estimates of β will be unreliable, but in these circumstances

standard test statistics contain the necessary information.

In Figure 2, the point o is anomolous, but since it occurs near the mean of X, no adverse leverage effects are inflicted on the slope estimate although the intercept will be affected. The source of this discrepant observation might be in X, Y or ε. If the latter, it could be indicative of heteroscedasticity or thick-tailed error distributions; clearly more such points are needed to analyze those problems further, but isolating the single point is constructive.

Figure 3 illustrates an instance of leverage where a gap arises between the main body of data and the outlier. While it constitutes a disproportionate amount of weight in the determination of β, it might be that benign third source of leverage mentioned above which supplies crucially useful information. Figure 4 is a more troublesome configuration that can arise in practice. In this situation the estimated regression slope is almost wholly determined by the extreme point. In its absence, the slope might be almost anything. Unless the extreme point is a crucial and valid piece of evidence (which of course depends on the research context), the researcher is likely to be highly suspicious of the estimate. Given the gap and configuration of the main body of data, the estimate surely has less than n-2 degrees of freedom; in fact it might appear that there are effectively two data points altogether, not n.

Finally, the leverage displayed in Figure 5 is a potential source of concern since o and/or • will heavily influence β but differently than the remaining data. Here is a case where deletion of data, perhaps less drastic downweighting, or model reformulation is clearly indicated.

Plots for Alternative Configurations of Data

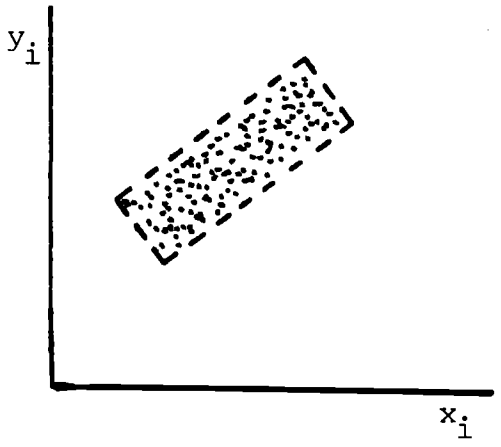


Figure 1

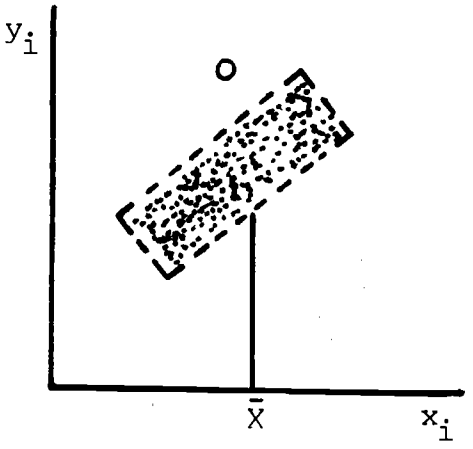


Figure 2

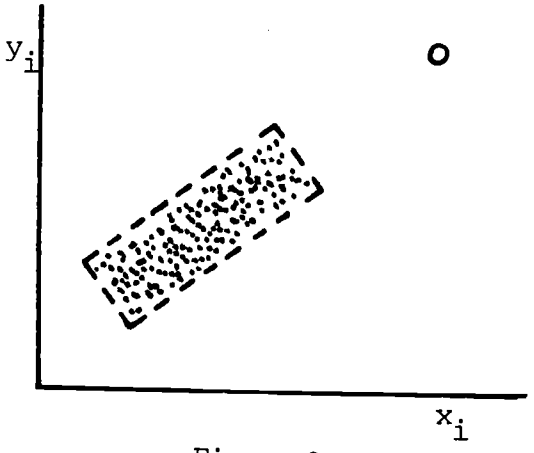


Figure 3

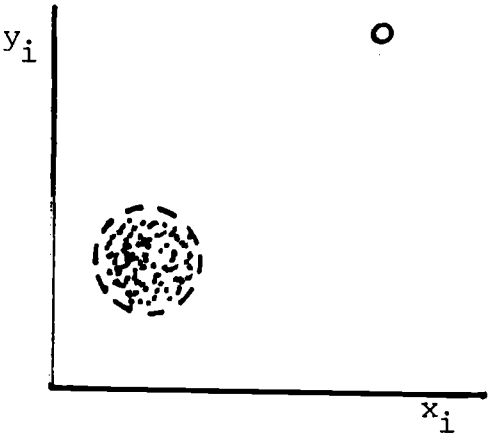


Figure 4

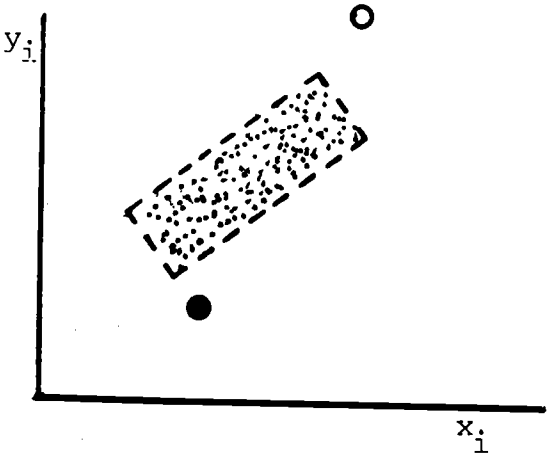


Figure 5

2.2 Residual Diagnostics

Traditionally the examination of functions of the residuals, $r_i = y_i - \hat{y}_i$, and especially large residuals, has been used to provide indications of suspect data that in turn may unduly affect regression results. It is best to have a scalar covariance matrix, so that detection of heteroscedasticity or autocorrelation (and later on, eliminating them) is desirable.

Approximate normality is another desirable property in terms of estimation efficiency and the ability to test hypotheses. Harmful departures from normality include pronounced skewness, multiple modes and thick-tailed error distributions. Even moderate departures from normality can noticeably impair estimation efficiency. At the same time, large outliers in error space will often be associated with modest-sized residuals in least squares estimates since the squared error criterion heavily weights extreme values.

It will often be difficult in practice to distinguish between heteroscedasticity and thick-tailed error distributions; to observe the former, a number of dependent variable values must be associated with (at least) several given configurations of explanatory variables. Otherwise, a few large residual outliers could have been generated by a thick-tailed error distribution or fragments from a heteroscedastic distribution.

Relevant diagnostics have three aspects, two of which examine the residuals and the third involving a change in error distribution assumptions. The first is simply a frequency distribution of the residuals. If there is evident visual skewness, multiple modes or a heavy tailed distribution, the graph will prove informative. It is interesting to note that economists often look at time plots of residuals but seldom at their frequency distribution.

The second is the normal probability plot, which displays the cumulative normal distribution as a straight line whose slope measures the standard deviation and whose intercept reflects the mean. Thus departures from normality of the cumulative residual plot will show up in noticeable departures from a straight line. Outliers will appear immediately at either end of the cumulative distribution.

Finally, Denby and Mallows [17] and Welsch [18] have suggested plotting the estimated coefficients and residuals as the error density or, equivalently, as the loss function (negative logarithm of the density) is changed. One family of loss functions has been suggested by Huber [1];

$$\rho(t) = \begin{cases} \frac{t^2}{2} & |t| \leq c \\ c|t| - \frac{c^2}{2} & |t| > c \end{cases}$$

which goes from least-squares ($c=\infty$) to least absolute residuals ($c=0$). This approach is attractive because of its relation to robust estimation [1], but requires considerable computation.

For diagnostic use the residuals can be modified in ways that will enhance our ability to detect problem data. We first note that the r_i do not have equal variances because if we let $H = X(X^TX)^{-1}X^T$, then

$$E[(Y-\hat{Y})(Y-\hat{Y})^T] = E[(I-H)YY^T(I-H)^T]$$

$$= (I-H)\, E(YY^T)(I-H) = \sigma^2(I-H)$$

since $(I-H)^2 = I-H$ and $(I-H)X = 0$. (See Theil [10] and Hoaglin and Welsch [13] for a more detailed discussion.) Thus

$$\text{var}\,(r_i) = \sigma^2\,(1-h_i) \qquad (2.2.1)$$

where h_i is the $i^{\underline{th}}$ diagonal element of H.

Consequently a number of authors [11] have suggested that instead of studying r_i, we should use the standardized residuals

$$r_{si} = r_i/s\sqrt{1-h_i} \tag{2.2.2}$$

where s^2 is the estimated error variance.

For diagnostic purposes we might want to go further and ask about the size of the residual corresponding to y_i when data point i has been omitted from the fit, since this corresponds to a simple perturbation of the data. That is, we base the fit on the remaining n-1 data points and then predict the value for y_i. This residual is

$$\tilde{r}_i = y_i - x_i \hat{\beta}_{(i)} \tag{2.2.3}$$

and has been studied in a different context by Allen [12]. Similarly $s^2_{(i)}$ is the estimated error variance for the "not i" fit, and the standard deviation of $\tilde{r}_i$ is estimated by $s_{(i)}\sqrt{1 + x_i(X^T_{(i)}X_{(i)})^{-1}x_i^T}$. We now define the studentized residual:

$$r_i^* = \frac{y_i - x_i \hat{\beta}_{(i)}}{s_{(i)}\sqrt{1 + x_i(X_{(i)}{}^T X_{(i)})^{-1} x_i^T}} \quad . \tag{2.2.4}$$

Since the numerator and denominator in (2.2.4) are independent, r_i^* has a t distribution with n-p-1 degreees of freedom. Thus we can readily assess the significance of any single studentized residual. (Of course, r_i^* and r_j^* will not be independent.) Perhaps even more useful for our purposes is the fact that

$$r_i^* = r_i/\left(s_{(i)}\sqrt{1-h_i}\right) \tag{2.2.5}$$

and

$$(n-p-1)s^2_{(i)} = (n-p)s^2 - \frac{r_i^2}{1-h_i} \qquad (2.2.6)$$

These results are proved easily by using the matrix identities in Appendix 1.

Therefore we think that a good way to examine residuals is to look at the studentized residuals, both because they have equal variances and because they are easily related to the t-distribution. However this does not tell the whole story, since some of the most influential data points can have relatively small studentized residuals (and very small r_i).

To illustrate with the simplest case, regression through the origin, we have

$$r_i = \sum_{j \neq i} x_j^2 \quad \frac{\hat{\beta} - \hat{\beta}_{(i)}}{x_i} \qquad (2.2.7)$$

$$\hat{\beta} - \hat{\beta}_{(i)} = x_i r_i / \sum_{j \neq i} x_j^2 \qquad (2.2.8)$$

where (i) denotes an estimate obtained by removing the $i^{\underline{th}}$ row (data point) from the computation. Thus the residuals are related to the change in the least-square estimate caused by deleting one row. But each contains different information since large values of $|\hat{\beta} - \hat{\beta}_{(i)}|$ can be associated with small $|r_i|$ and vice versa. Therefore we are lead to consider row deletion as an important diagnostic tool, to be treated on at least an equal footing with the analysis of residuals.

For multivariate linear regression (2.2.8) becomes

$$\hat{\beta} - \hat{\beta}_{(i)} = (X^T X)^{-1} x_i^T r_i / (1-h_i) \qquad (2.2.9)$$

where the h_i are the diagonal elements of H, the least-squares projection matrix defined earlier. We will call this the "hat" matrix since

$$HY = \hat{Y} = X\hat{\beta} \quad . \tag{2.2.10}$$

Clearly the hat matrix plays a crucial role not only in the studentized residuals but also in row deletion and other diagnostic tools. We now develop some important results (based on the discussion in Hoaglin and Welsch [13]) relating to this matrix.

2.3 The Hat Matrix

Geometrically $\hat{Y}$ is the projection of Y onto the p-dimensional subspace of n-space spanned by the columns of X. The element h_{ij} of H has a direct interpretation as the amount of leverage or influence exerted on $\hat{y}_i$ by y_j. Thus a look at the hat matrix can reveal sensitive points in the X space, points at which the value of y has a large impact on the fit.

The influence of the response value y_i on the fit is most directly reflected in its leverage on the corresponding fitted value $\hat{y}_i$, and this is precisely the information contained in h_i, the corresponding diagonal element of the hat matrix. When there are two or fewer explanatory variables scatter plots will quickly reveal any x-outliers, and it is not hard to verify that they have relatively large h_i values. When $p > 2$, scatter plots may not reveal "multivariate outliers," which are separated in p-space from the bulk of the x-points but do not appear as outliers in a plot of any single explanatory variable or pair of them

yet will be revealed by an examination of H. Looking at the diagonal elements of H is not absolutely conclusive but provides a basic starting point. Even if there were no hidden multivariate outliers, computing and examining H (especially the h_i) is usually less trouble than looking at all possible scatter plots.

As a projection matrix, H is symmetric and idempotent ($H^2 = H$). Thus we can write

$$h_{ii} = \sum_{j=1}^{n} h_{ij}^2 = h_{ii}^2 + \sum_{j \neq i} h_{ij}^2 \qquad (2.3.1)$$

and it is clear that $0 \leq h_{ii} \leq 1$. These limits are useful in understanding and iterpreting $h_i (\equiv h_{ii})$, but they do not yet tell us when h_i is "large". It is easy to show, however, that the eigenvalues of a projection matrix are either 0 or 1 and that the number of non-zero eigenvalues is equal to the rank of the matrix. In this case rank (H) = rank (X) = p and hence trace H = p, that is,

$$\sum_{i=1}^{n} h_i = p \quad . \qquad (2.3.2)$$

The average size of a diagonal element, then, is p/n. If we were designing an experiment a desirable goal would be to have all the data points be about equally influential or all h_i nearly equal. Since the X data is given to us and we cannot design our experiment to keep the h_i equal, we will follow [13] and say that h_i is a <u>leverage</u> point if $h_i > 2p/n$. We shall see later that leverage points can be both harmful and helpful.

The quantity 2p/n has worked well in practice and there is some theoretical justification for its use. When the explanatory variables are multivariate Gaussian it is possible to compute the exact distribution of certain functions of the h_i. Let $\tilde{X}$ denote the $n\times(p-1)$ matrix obtained by centering the explanatory variables. Now

$$\hat{Y} - \bar{Y} = HY - \bar{Y} = \tilde{H}Y \tag{2.3.3}$$

and thus the diagonal elements of the centered hat matrix are

$$\tilde{h}_i = h_i - \frac{1}{n} \quad . \tag{2.3.4}$$

Let $\tilde{X}_{(i)}$ denote $\tilde{X}$ with the i^{th} row removed and $\hat{X}_{(i)}$ denote the centered version of $X_{(i)}$, i.e. means based on all but the i^{th} observation have been subtracted out. Finally note that

$$x_i-\bar{x} = \frac{n-1}{n}(x_i-\bar{x}_{(i)}) \tag{2.3.5}$$

and

$$\bar{x}_{(i)}-\bar{x} = \frac{1}{n}(\bar{x}_{(i)}-x_i). \tag{2.3.6}$$

Using (A1.1) and (2.3.5)

$$\tilde{h}_i = \frac{\gamma}{1+\gamma}$$

where $\gamma = \left(\frac{n-1}{n}\right)^2 (x_i-\bar{x}_{(i)})(\tilde{X}_{(i)}^T \tilde{X}_{(i)})^{-1} (x_i-\bar{x}_{(i)})^T$.

Again using (A1.1) and (2.3.6)

$$\gamma = \left(\frac{n-1}{n}\right)^2 \frac{\alpha}{1+\frac{(n-1)}{n^2}\alpha}$$

where

$$\alpha = (x_i - \bar{x}_{(i)}) \, (\hat{X}_{(i)}^T \hat{X}_{(i)})^{-1} \, (x_i - \bar{x}_{(i)})^T .$$

The distribution of $(n-2)\alpha$ is well known since it is the Mahalanobis distance between observation i and the mean of the remaining observations [19, p. 480]. Thus

$$\alpha \sim \frac{n(p-1)}{(n-1)(n-p)} \, F_{p-1,n-p} . \tag{2.3.7}$$

Reversing the above algebraic manipulations we obtain

$$\tilde{h}_i = \frac{n-1}{n} \left[\frac{\alpha}{\frac{n}{n-1} + \alpha} \right]$$

and

$$h_i = \frac{(n-1)\alpha + 1}{(n-1)\alpha + n} .$$

Solving for α gives

$$\alpha = \frac{n}{n-1} \left(\frac{h_i - 1/n}{1-h_i} \right)$$

and from (2.3.7)

$$\frac{h_i - 1/n}{1-h_i} = \frac{n-1}{n} \, \alpha \sim \frac{p-1}{n-p} \, F_{p-1,n-p} . \tag{2.3.8}$$

For moderate p and moderate n the 95% point for F is near 2. Therefore, a cut-off point would be

$$h_i > \frac{2(p-1) + \frac{n-p}{n}}{n+p-2} \tag{2.3.9}$$

which is approximated by $2p/n$.

From equation (2.3.1) we can see that whenever $h_i = 0$ or 1, we have $h_{ij} = 0$ for all $j \neq i$. These two extreme cases can be interpreted as follows. If $h_i = 0$, then $\hat{y}_i$ must be fixed at zero - it is not affected by y_i or by any other y_j. A point with $x_i = 0$ when the model is a straight line through the origin provides a simple example.

When $h_i = 1$, we have $\hat{y}_i = y_i$ - the model always fits this data value exactly. This is equivalent to saying that, in some coordinate system, one parameter is determined completely by y_i or, in effect, dedicated to one data point. The following theorems are proved in appendix 3.

Theorem: If $h_i = 1$, there exists a nonsingular transformation, T, such that the least-squares estimates of $\alpha = T^{-1}\beta$ have the following properties: $\hat{\alpha}_1 = y_i$ and $\{\hat{\alpha}_j\}_{j=2}^{p}$ do not depend on y_i.

Theorem: If X is nonsingular, then

$$\det(X_{(i)}^T X_{(i)}) = (1-h_i)\det(X^T X) \quad . \tag{2.3.10}$$

Clearly when $h_i = 1$ the new matrix $X_{(i)}$ formed by deleting a row is singular and we cannot obtain the usual least-squares estimates. This is extreme leverage and does not often occur in practice.

To complete our discussion of the hat matrix we give a few simple examples. For the sample mean all elements of H are $1/n$. Here $p = 1$ and each $h_i = p/n$, the perfectly balanced case.

For a straight line through the origin

$$h_{ij} = x_i x_j / \sum_{k=1}^{n} x_k^2 \tag{2.3.11}$$

and clearly $\sum_{i=1}^{n} h_i = p = 1$.

Simple linear regression is slightly more complicated but a few steps of algebra give

$$h_{ij} = \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{\sum_{k=1}^{n} (x_k - \bar{x})^2} \tag{2.3.12}$$

and $\sum_{i=1}^{n} h_i = 2$. We can see from (2.3.12) how x-values far from $\bar{x}$ will lead to large values of h_i. It is this idea in the multivariate case that we attempt to capture by looking at elements of the hat matrix.

2.4 Row Deletion Diagnostics

We now return to the basic formula

$$\hat{\beta} - \hat{\beta}_{(i)} = (X^T X)^{-1} x_i \, r_i/(1-h_i). \tag{2.4.1}$$

Since the variability of $\hat{\beta}_j$ is measured by $s((X^T X)^{-1}_{jj})^{\frac{1}{2}}$, a more useful measure of change is

$$\text{DFBETAS}_{ij} = \frac{(\beta - \beta_{(i)})_j}{s_{(i)}\sqrt{(X^T X)^{-1}_{jj}}} \tag{2.4.2}$$

where we have replaced s by $s_{(i)}$ in order to make the denominator stochastically independent of the numerator in the Gaussian case. To provide a summary of the relative coefficient changes we suggest

$$NDFBETAS_i = \sqrt{\frac{n-p}{p} \sum_{j=1}^{p} DFBETAS_{ij}^2} \quad . \tag{2.4.3}$$

The term $\frac{n-p}{p}$ has been incorporated to make NDFBETAS more comparable across data sets which may have different values of p and n. This normalizing value was chosen because when X is an orthogonal matrix (but not necessarily orthonormal)

$$DFBETAS_{ij} = \frac{x_{ij}\, r_i^*}{\sum_{t=1}^{n} x_{tj}^2 \sqrt{1-h_i}}$$

and

$$\sum_j DFBETAS_{ij}^2 = \frac{h_i}{1-h_i}\, r_i^{*\,2} \quad .$$

Since the average value of $h_i = p/n$, a rough average value for $h_i/(1-h_i)$ is $p/(n-p)$. Clearly (2.4.3) could be modified to reflect the fact that some coefficients may be more important than others to the model builder (e.g., including only the main estimates of interest).

Another obvious row deletion diagnostic is the change in fit

$$DFFIT_i = x_i(\hat{\beta}-\hat{\beta}_{(i)}) = \frac{h_i}{1-h_i}\, r_i \quad . \tag{2.4.4}$$

If we scale this by dividing by $s_{(i)}\sqrt{h_i}$ we have

$$\frac{\sqrt{h_i}}{\sqrt{1-h_i}}\, r_i^* \quad . \tag{2.4.5}$$

For across data set normalization we will multiply by $\sqrt{n-p/p}$ to obtain

$$\text{DFFITS}_i = \sqrt{\left(\frac{n-p}{p}\right)\left(\frac{h_i}{1-h_i}\right)\ r_i^{*\,2}} \qquad (2.4.6)$$

A measure similar to this has been suggested by Cook [14].

Clearly DFFITS and NDFBETAS agree in an orthogonal coordinate system. When orthogonality does not hold these two measures provide somewhat different information. Since we tend to emphasize coefficients, our preference is for NDFBETAS.

Deciding when a difference like $|(\hat{\beta} - \hat{\beta}_{(i)})_j|$ or other diagnostic statistic is large will depend, in part, on how this information is being used. For example, large changes in coefficients that are not of particular interest might not overly upset the model builder while a change in an important coefficient may cause considerable concern even though the change is small relative to traditional estimation error.

We have used two approaches to measure the size of changes caused by row deletion. The first, called external comparison, generally uses measures associated with the quantity whose changes are being studied. For example, the standard error of a particular coefficient $\hat{\beta}_j$ would be used with $(\hat{\beta} - \hat{\beta}_{(i)})_j$.

The second method, called internal comparison, treats each set of diagnostic values (e.g., $\{(\hat{\beta} - \hat{\beta}_{(i)})_j\}_{i=1}^{n}$) as a single data series and then finds, for example, the standard deviation of this series as a measure of relative size. As we have noted, all of the diagnostic measures we have discussed so far are functions of $r_i/\sqrt{1-h_i}$ and in view of our discussion of studentized residuals, it is natural to divide this by $s_{(i)}$ to achieve a reasonable scaling before making plots, etc.

Once $s_{(i)}$ has been used, the temptation arises to try to perform formal statistical tests because we know the distribution of r_i^*. In our opinion this is not a very promising procedure because it puts too much emphasis on residuals (although looking at studentized residuals is better than using the raw residuals). We prefer to use external or internal comparison to make decisions about which data points deserve further attention except, of course, when we are looking specifically at the studentized residuals as we did earlier. Using any Gaussian distributional theory depends on the appropriateness of the Gaussian error distribution - a topic we will return to later.

2.5 Regression Statistics

Most users of statistics realize that estimates like $\hat{\beta}$ should have some measure of variability associated with them. It is less often realized that regression statistics like t, R^2 and F should also be thought of as having a variability associated with them.

One way to assess this variability is to examine the effects of row deletion on these regression statistics. We have focused on three:

$$\Delta TSTAT_j = \frac{\hat{\beta}_j}{s.e.(\hat{\beta}_j)} - \frac{(\hat{\beta}_{(i)})_j}{s.e.(\hat{\beta}_{(i)})_j}$$

$$\Delta FSTAT = F(\text{all } \beta = 0) - F_{(i)}(\text{all } \beta = 0)$$

$$\Delta R^2 = R^2 - R^2_{(i)} .$$

Again we should ask when a difference is large enough to merit attention. For external comparison we would compare to the standard deviation of t, F, or R^2:

Statistic	Standard Deviation
t	$\left(\frac{n-p}{n-p-2}\right)^{1/2}$
F	$\left(\frac{2(n-p)^2(n-2)}{p(n-p-2)(n-p-4)}\right)^{1/2}$
R^2	$\left(\frac{(p-1)^2}{(p+n-2)^2(p+n-1)}\right)^{1/2}$.

However, we tend to view internal comparison as more appropriate for regression statistics.

Studying the changes in regression statistics is a good second order diagnostic tool because if a row appears to be overly influential on other grounds, an examination of the regression statistics will show if the conclusions of hypothesis testing would be affected.

There is, of course, room for misuse of this procedure. Data points could be removed solely on the basis of their ability (when removed) to increase R^2 or some other measure. While this danger exists we feel that it is often offset by the ability to study changes in regression statistics caused by row deletion. Again we want to emphasize that changes in regression statistics should not be used as a primary diagnostic tool.

2.6 Influence and Variance Decomposition

We now would like to consider perturbing our assumptions in a new way. Consider the standard regression model (1.4) but with $\text{var}(\varepsilon_i)$ replaced by σ^2/w_i for just the i^{th} data point. In words, we are perturbing the homoscedasticity assumption for this one data point.

In appendix 2 we show that

$$\frac{\partial \hat{\beta}_{w_i}}{\partial w_i} = \frac{(X^TX)^{-1}x_i^Tr_i}{(1-(1-w_i)h_i)^2} \tag{2.6.1}$$

and it follows that

$$\left.\frac{\partial \hat{\beta}_{w_i}}{\partial w_i}\right|_{w_i=1} = (X^TX)^{-1}x_i^Tr_i \tag{2.6.2}$$

$$\left.\frac{\partial \hat{\beta}_{w_i}}{\partial w_i}\right|_{w_i=0} = \frac{(X^TX)^{-1}x_i^Tr_i}{(1-h_i)^2} = \frac{\hat{\beta}-\hat{\beta}_{(i)}}{1-h_i} \; . \tag{2.6.3}$$

Equation (2.6.2) tells us about infinitesimal changes in $\hat{\beta}$ caused by small changes in w_i about the value 1 and similarly for (2.6.3). From the mean value theorem we know that

$$\hat{\beta}-\hat{\beta}_{(i)} = \left.\frac{\partial \hat{\beta}_{w_i}}{\partial w_i}\right|_{\delta} \tag{2.6.4}$$

where δ is between 0 and 1. Any one of (2.6.2), (2.6.3) or (2.6.4) can be used for diagnostic purposes. We have chosen to emphasize $\hat{\beta}-\hat{\beta}_{(i)}$ because of its intuitive appeal and the fact that it is a compromise between (2.6.2) and (2.6.3).

Formula (2.6.2) can also be considered as a function which represents the influence of the i^{th} data point and can be linked to the theory of robust estimation [15] and the jackknife [16].

If we let

$$s^2_{w_i} = \frac{1}{n-p} \sum_{t=1}^{n} w_t (y_t - x_t \hat{\beta}_{w_i})^2 \tag{2.6.5}$$

and

$$W = \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & w_i & \\ & & & \ddots \\ & & & & 1 \end{bmatrix} \tag{2.6.6}$$

then in appendix 2 we show that

$$\frac{\partial}{\partial w_i} \left[s^2_{w_i} (X^T W X)^{-1} \right] \Big|_{w_i=1} =$$

$$\frac{r_i^2}{n-p} (X^T X)^{-1} - s^2 (X^T X)^{-1} x_i^T x_i (X^T X)^{-1} . \tag{2.6.7}$$

Since we would like to remove scale we define

$$DBVARS_{ij} = \frac{r_i^2}{(n-p)s^2} - \frac{\left[(X^T X)^{-1} x_i^T x_i (X^T X)^{-1} \right]_{jj}}{(X^T X)^{-1}_{jj}} \tag{2.6.8}$$

as the scaled infinitesimal change in the variance of $\hat{\beta}_j$. As a summary measure over all of the coefficients we use

$$NDBVARS_i = \frac{n}{p} \sum_{j=1}^{p} |BVARS_{ij}| . \tag{2.6.9}$$

where the n/p term is used to improve comparability across data sets.

If we used row deletion instead of derivatives, our basic measure would be

$$DFBVARS_{ij} = \frac{s^2 (X^TX)^{-1}_{jj} - s^2_{(i)}(X^T_{(i)}X_{(i)})^{-1}_{jj}}{s^2 (X^TX)^{-1}_{jj}} \qquad (2.6.10)$$

with summary measure

$$NDFBVARS_i = \frac{(n-p)}{p} \sum_{j=1}^{p} |DFBVARS_{ij}| \ . \qquad (2.6.11)$$

The measures so far discussed in this section include both the explanatory variables and the response. If we wish to examine the X-matrix only, the second, part of (2.6.8) provides a good way to do this. We notice that

$$\sum_{i=1}^{n} (X^TX)^{-1} x_i^T x_i (X^TX)^{-1} = (X^TX)^{-1}$$

and define

$$BETAVRD_{ij} = \frac{\left[(X^TX)^{-1} x_i^T x_i (X^TX)^{-1}\right]_{jj}}{(X^TX)^{-1}_{jj}}$$

with summary measure

$$NBETAVRD_i = \sum_{j=1}^{p} BETAVRD_{ij}$$

These measures provide a way to decompose the cross products matrix with respect to the individual observations.

Again it is useful to look at the orthogonal X case. When orthogonality holds

$$\text{BETAVRD}_{ij} = \frac{x_{ij}^2}{\sum_{i=1}^{n} x_{ij}^2}$$

and

$$\text{NBETAVRD}_i = h_i .$$

Since h_i has a strong intuitive appeal it may be a better summary value even when orthogonality does not hold. We have chosen not to multiply NBETAVRD by n/p (the average value for h_i), so it is not useful across data sets.

If we examine the formula for DFBVARS we see that this quality could be positive or negative. As we might expect, in some cases downweighting a data point can improve our estimate of the variance of a coefficient. (Downweighting corresponds to placing a minus sign in front of DFBVARS.) One of the best ways to examine the tradeoffs of DFBETAS and DFBVARS (or BETAVRD) is to make a scatter plot. A high leverage point with small values of DFBETAS may be a "good" observation because it is helping to reduce the variance of certain coefficients. The setting aside of all high leverage points is generally not an efficient procedure because it fails to take account of the response data.

2.7 More Than One Row at a Time

It is natural to ask if there might be groups of leverage points that we are failing to diagnose because we are only looking at one row at a time. There are easily constructed examples where this can happen.

One approach is to proceed sequentially - remove the "worst" leverage point (based perhaps on both NDFBETAS and NBETAVRD), reexamine the diagnostic measures and remove the next "worst" observation, etc. This does not fully

cope with the problem of groups of leverage points and just as stepwise regression can be troublesome, so can sequential row deletion.

A straightforward induction argument shows that

$$\delta_{ij} - h_{ij}(k_1,k_2,\ldots k_t)$$

$$= \frac{\det (I-H)_{i,k_1,k_2,\ldots k_t;j,k_1,k_2,\ldots k_t}}{\det (I-H)_{k_1,k_2,\ldots k_t}}$$

where H is the hat matrix for all of the data, $h_{ij}(k_1,\ldots,k_t)$ denotes the hat matrix for a regression with rows $k_1,\ldots k_t$ removed and the subscripts on I-H denote a submatrix formed by taking those rows and columns of I-H.

Even though all of these differences are based on H, multiple row deletion will involve large amounts of computation. It is instructive to note that

$$1-h_i(k) = \frac{(1-h_i)(1-h_k) - h_{ik}^2}{1-h_k}$$

$$= (1-h_i) \left[1 - \frac{h_{ik}^2}{(1-h_k)(1-h_i)} \right]$$

$$= (1-h_i) \quad [1 - \text{cor } (r_i,r_k)]$$

The term cor (r_i,r_k) also appears when more rows are deleted and, in place of looking at all possible subsets of rows, an examination of the correlation matrix of the residuals for large correlations has provided useful clues to groups of leverage points. This requires knowing the off-diagonal values of H and therefore increases computational cost and perhaps storage requirements.

2.8 Interface with Robust and Ridge Regression

It is natural to ask how the above diagnostics could or should be used with some of the newer estimation methods like robust and ridge regression. The first question is whether we should do diagnostics or robust or ridge first. There is no clear answer, but some sort of iterative procedure is probably called for.

However, it is possible to perform regression diagnostics after using either a robust procedure or a ridge procedure. In the robust case we can make use of weights

$$w_i = \rho'\left[\frac{(y_i - x_i\hat{\beta}_R)}{\hat{s}_R}\right] \qquad (2.8.1)$$

where ρ is the robust loss function, $\hat{\beta}_R$ are the robust estimates of β and $\hat{s}_R$ is a robust estimate of the scale of the residuals, $y_i - x_i\hat{\beta}_R$. (A complete discussion of weights is contained in [20].) We now modify the data by forming a diagonal matrix of weights, W, and using $\sqrt{W}Y$, $\sqrt{W}X$. This revised data is then the input to regression diagnostics. If the robust estimation procedure has been allowed to converge

$$\hat{\beta}_w = (X^TWX)^{-1}X^TWY$$

will be close to $\hat{\beta}_R$ and our procedures will accurately reflect what would happen to $\hat{\beta}_R$ locally. Of course they do not reflect what would happen if a data point were deleted and then robust estimation applied.

The ridge estimator [21] is given by

$$\hat{\beta}_{RD} = (X^TX + kI)^{-1}X^TY\ . \qquad (2.8.2)$$

There are many generalizations but most will fit into the following framework. We assume that k has been chosen by some means such as those listed

in [21]. Then we form

$$X_A = \begin{bmatrix} X \\ \sqrt{k}\, I_{p\times p} \end{bmatrix}, \quad Y_A = \begin{bmatrix} Y \\ 0_p \end{bmatrix}$$

where 0_p is a p×1 vector of zeros (prior values times $\sqrt{k}$ in more general cases). So we now have "new" data X_A and Y_A with n×p rows. Clearly

$$\hat{\beta}_{RD} = (X_A^T X_A)^{-1} X_A^T Y_A . \qquad (2.8.3)$$

We now perform regression diagnostics using X_A and Y_A. When we delete a row with index n+j > n, it is equivalent to saying we do not want to "shrink" that parameter estimate toward zero (or its prior). In the Bayesian context dropping such a row is like setting the prior precision of β_j to zero. Plots of DFBETAS would then show the effects of such a process by looking at those DFBETAS values for index greater than n.

We can do some diagnostics to decide if a ridge estimator is warranted. If we differentiate (2.8.2) with respect to k, then

$$\frac{\partial \hat{\beta}_{RD}}{\partial k} = (X^T X + kI)^{-1} \hat{\beta}_{RD} \qquad (2.8.4)$$

and

$$\left.\frac{\partial \hat{\beta}_{RD}}{\partial k}\right|_0 = (X^T X)^{-1} \hat{\beta} \qquad (2.8.5)$$

Thus (2.8.5) provides information about infinitesimal charges about k=0. If $X^T X$ were diagonal then (2.8.5) has components $\hat{\beta}_j/\lambda_j$ where λ_j are the eigenvalues. So $\hat{\beta}_j$ large and/or λ_j small would lead to a large value of the derivative. Since the ridge estimator depends heavily on the scaling of the explanatory variables, so does (2.8.4) and we recommend scaling before using this diagnostic measure.

When diagnostics have been completed a few observations may be suspect. The rows can then be set aside and a new robust or ridge estimate computed. Diagnostics can then be applied again. There are obvious limits of time and money but we think that two passes through this process will often be worthwhile.

2.9 An Example: An Inter-Country Life Cycle Savings Function

Arlie Sterling of MIT has made available to us data he has collected on fifty countries in order to undertake a cross-sectional study of the life cycle saving hypothesis. The savings ratio (aggregate personal saving divided by disposal income) is explained by per capita disposable income, the percentage rate of change in per capita disposable income and two population variables: per cent less that 15 years old and per cent over 75 years old. The data are averaged over the decade 1960-1970 to remove the business cycle or other short-term fluctuations.

According to the life cycle hypothesis, savings rates should be negatively affected if non-members of the labor force constitute a large part of the population. Income is not expected to be important since age distribution and the rate of income growth constitute the core of life cycle savings behavior. The regression equation and variable definitions are then:

$$SR_i = COEF.1 + COEF.2*POP15_i + COEF.3*POP75_i + COEF.4*INC_i + COEF.5*INGRO_i \qquad (2.9.1)$$

SR_i = the average aggregate personal savings rate in country i from 1960-1970

$POP15_i$ = the average % of the population under 15 years of age from 1960-1970

$POP75_i$ = the average % of the population over 75 years of age from 1960-1970

INC_i = the average level of real per capita disposable income in country i from 1960-1970 measured in U.S. dollars

$INGRO_i$ = the average % growth rate of INC_i from 1960-1970.

A full list of countries, together with their numerical designation, appears in Exhibit 1, and the data are in Exhibit 2. It is evident that a wide geographic area and span of economic development are included. It is also plausible to suppose that the quality of the underlying data is highly variable. With these obvious caveats, the LS estimates of (2.9.1) are shown in Exhibit 3. To comment briefly, the R^2 is not uncharacteristically low for cross-sections, the population variables have correct negative signs - COEF 3 has a small t statistic but COEF 2 does not - income is statistically insignificant, while income growth reflected in COEF 5 is significant at the 5 per cent level and has a positive influence on the savings rate as it should. Broadly speaking, these results are consistent with the life cycle hypothesis.

The remainder of this section will be a guided tour through some of the diagnostics discussed previously. The computations were performed using SENSSYS (acronym for sensitivity system), a TROLL experimental subsystem for regression diagnostics. Orthogonal decompositions are used in the least-squares regression computations and this makes it possible to get all of the diagnostic measures in addition to the usual LS results in less than twice the computer time for the LS results alone.

David Jones and Steve Peters of the NBER Computer Research Center have programmed SENSSYS. Both have actively participated in analytical and empirical aspects of the research.

Only a selection of plots and diagnostics will be shown for two reasons. One is that to provide the full battery of plots would be excessively tedious; however, the missing plots and tables are readily obtainable. The other reason is that we found these diagnostics to be among the more instructive from examination of this and several other problems.

2.9.1 Residuals

The first plot, Exhibit 4, is a normal probability plot. Departure from a fitted line (which represents a particular Gaussian distribution with mean equal to the intercept and standard deviation equal to the slope) is not substantial in the main body of the data for these studentized residuals, but Zambia (46) is an extreme residual which departs from the line. Different information, an index plot of the r_i, appears in Exhibit 5 which reveals not only Zambia, but possibly Chile (7) as well to be an outlier; each exceeds 2.5 times the standard error.

2.9.2 Leverage and Diagonal Hat Matrix Entries

Exhibit 6 plots the h_i which, as diagnonals of the hat matrix, are indicative of leverage points. Most of the h_i are small, but two stand out sharply: Libya (49) and the United States (44). Two others, Japan (23) and Ireland (21) exceed the $2p/n = .20$ criterion (which happens to be equal to the 95% significance level based on the F distribution), but just barely. Deciding whether or not leverage is potentially detrimental depends on what happens elsewhere in the diagnostic analysis, although it should be recalled that it is values near unity that pose the most severe problems, which has not happened here.

2.9.3 Coefficient Perturbation

An overview of the effects of individual row deletion (see Exhibit 7) is based on (2.4.3) NDFBETAS, the square root of the scaled sum of the squared differences between the full data set and row deleted coefficients. The measure used is scaled approximately as the t distribution so that values greater than 2 are a potential source of concern. Two countries that also showed up as possible high leverage candidates, Libya (49) and Japan (23), also seem to have a heavy influence on the coefficients while Ireland (21), a marginal high leverage candidate, is also a marginal candidate for influencing coefficient behavior. Individual plots of DFBETAS (2.4.2) follow next, from which the following table has been constructed based on an examination of Exhibits 8-11.

Noticeably Large Effects on $\hat{\beta}_j$ from Row Deletion

Population <15	Population >75	Income	Income Growth
Japan (23)	Ireland (21)		Libya (49)
	Japan (23)		Japan (23)

The countries that stand out in the individual coefficients are perhaps, not surprisingly, the two that appeared in the overall measure. Ireland, in addition, appears once. Except on the income variable, the comparatively large values are just about one LS standard error for each particular coefficient.

2.9.4 Variation in Coefficient Standard Errors

Exhibit 12 is a summary measure of coefficient standard error variations as a consequence of row deletions, designated as NDFBVARS in (2.6.9). Since these standard errors involve both error variance and elements from $(X^TX)^{-1}$,

large values indicate simultaneous or individual extremes in residuals or multivariate outliers in the X matrix. These quite numerous candidates include:

Index	Country
7	Chile
21	Ireland
23	Japan
37	Southern Rhodesia
44	United States
46	Zambia
49	Libya

Of these seven countries, six appeared previously, while the only new candidate is Southern Rhodesia. Libya had both high leverage and large coefficient changes, Ireland and Japan had noticeable coefficient changes, while Chile and Zambia possess large residuals. Thus this particular diagnostic may have some use as a comprehensive measure.

Plots for percent changes in the individual coefficient standard errors are shown in Exhibits 13-16. Large individual changes (here taken to be in excess of 25%) appear for the United States with a 47% change for the income variable, while the deletion of Libya increases the standard error for the same variable by nearly 85%.

2.9.5 Change in Fit

The standardized change in fit, DFFITS (2.4.6), with a row deleted, while similar in algebraic structure to coefficient change, conveys somewhat different information of general interest with specific applications in a time series context. DFFITS can be viewed in some theoretical cases as having a t distribution so that extremes of concern show up for values in excess of 2. In Exhibit 17 three countries that surfaced previously reappear: Japan (23), Zambia (46) and Libya (49). When coefficient changes

alone are considered as shown in Exhibit 7, Zambia did not appear, while Ireland (21) did. Thus somewhat different information is contained in each.

2.9.6 A Provisional Summary

It is now desirable to bring together the information that has been assembled thus far, to see what it all adds up to. One useful summary plot is shown in Exhibit 18, which plots the summary measure of $\hat{\beta} - \hat{\beta}_{(i)}$, NDFBETAS against the corresponding hat matrix diagonal, h_i.

The first point which emerges is that Japan (23) and Libya (49) have both high leverage and a significant influence on the estimated parameters. This is reason enough to view them as serious problems. (After the analysis had reached this point, we were informed by Arlie Sterling that a data error had been discovered for Japan. When corrected, he tells us that the revised data is more similar to the majority of countries. These diagnostics have thus "proven their worth" in bad data detection in a modest way. Second, Ireland is an in-between case, with moderately large leverage and a somewhat disproportionate impact on the coefficient estimates.

Third, the United States has high leverage combined with only meager differential effect on the estimated coefficients. Thus leverage in this instance can be viewed as neutral or beneficial. It is important to note that not all leverage points cause large changes in $\hat{\beta}$.

Exhibit 19 plots the summary of coefficient change, NDFBETAS against the studentized residuals and visually drives home the point that large residuals do not necessarily coincide with large changes in coefficients; all of the large changes in coefficients are associated with standardized residuals less than 2. Thus residual analysis alone is not a sufficient diagnostic tool.

Another summary plot, that of change in coefficient standard error, NDFBVARS against leverage as measured by h_i in Exhibit 20 indicates the close anticipated association between leverage and estimated parameter variability. This is clearly shown by the diagonal line composed of (21) Ireland, (23) Japan, (44) United States and (49) Libya. But residuals also can have a large and separate influence, as evidenced by the low leverage, high standard error changes for (7) Chile and (46) Zambia.

A final summary plot, Exhibit 21 of NDFBETAS against NDFBVARS, is revealing in that all of the points noted outside the cutoff points (3,2) have been spotted in the previous diagnostics as worth another look for one reason or another. Thus about 15% of the observations have been flagged, not an excessive fraction for many data sets.

2.9.7 One Further Step

Since Libya (49) is clearly an extreme and probably deleterious influence on the original regression, a reasonable next step is to eliminate it to find out whether its presence has masked other problems or not. Exhibit 22 plots the h_i when Libya (49) has been excluded in the data set. There is only one noticeable difference since Ireland (21), Japan (23) and the United States (44) remain high leverage points. Southern Rhodesia (37) now appears as a marginally significant leverage point, whereas it had previously been just below the cutoff. The only really new fact is that Jamaica (47) now appears as a prominent leverage point.

Jamaica has furthermore now become a source of parameter influence which is perhaps most effectively observed in the recalculation of scaled parameter changes, NDFBETAS, in Exhibit 23 which reveals Jamaica as the single largest source of overall coefficient variation.

This illustrates the proposition that perverse extreme points can mask the impact of still other perverse points. Yet the original analysis did contain most of the pertinent information about exceptional data behavior. The correlation matrix of the residuals discussed in Section 2.7 provided a clue, since the squared correlation between (47) and (49) was .173, the hightest value. It is nevertheless a prudent step to reanalyze the data with suspect points removed, to ascertain whether one or more extreme or suspect data points have obscured or dominated others.

2.10 Final Comments

The question naturally arises as to whether the approach we have taken in detection of outliers is more effective than simply examining each individual column of the data to look for detached observations. We believe the answer is yes. Detached outliers did appear in column 5 (INGRO) of the X matrix for Libya (49) and Jamaica (47), but not elsewhere. Libya, of course, was "the villain of the piece" in the prior analysis. But leverage points for numerous other countries were revealed by row deletion diagnostics, while Jamaica, as matters turned out, was not a particularly troublesome data point. In addition we discussed how various leverage points affected our output - coefficients, fit, or both. So we conclude at this early stage of our investigation, that these new procedures have merit in uncovering discrepant data that is not possible with a high degree of confidence by just looking at the raw data.

References

[1] Huber, P.J., "Robust Regression: Asymptotics, Conjectures and Monte Carlo," Annals of Statistics, 1 (1973), pp. 799-821.

[2] Belsley, David A., "Multicollinearity: Diagnosing its Presence and Assessing the Potential Damage it Causes Least-Squares Estimation," Working Paper 154, National Bureau of Economic Research, Computer Research Center for Economics and Management Science, October 1976, Cambridge, Mass.

[3] Silvey, S.D., "Multicollinearity and Imprecise Estimation," Journal of the Royal Statistical Society, Series B, Vol. 31, 1969, pp. 539-552.

[4] Theil, H., Linear Aggregation of Economic Relations, North Holland, Amsterdam, 1954.

[5] Chow, Gregory C., "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," Econometrica, Vol. 28, 1960, pp. 591-605.

[6] Fisher, Franklin M., "Tests of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note," Econometrica, Vol. 38, 1970, pp. 361-366.

[7] Goldfeld, Stephen M. and Richard E. Quandt, "The Estimation of Structural Shifts by Switching Regression," Annals of Economic and Social Measurement, Vol. 2, No. 4, 1973, pp. 475-485.

[8] Quandt, Richard E., "A New Approach to Estimating Switching Regressions," Journal of the American Statistical Association, Vol. 67, 1972, pp. 306-310.

[9] Brown, R.L., J. Durbin and J.M. Evans, "Techniques for Testing the Constancy of Regression Relationships," Journal of the Royal Statistical Society, Series B, Vol. 37, No. 2, 1975, pp. 149-163.

[10] Theil, H., Principles of Econometrics, John Wiley and Sons, New York, 1971, pp. 193-195.

[11] Anscombe, F.J. and Tukey, J.W., "The Examination and Analysis of Residuals," Technometrics, 5 (1963), pp. 141-160.

[12] Allen, David M., "The Relationship Between Variable Selection and Data Augmentation and a Method for Prediction," Technometrics, 16 (1974), pp. 125-127.

[13] Hoaglin, D.C. and Welsch, R.E., "The Hat Matrix in Regression and Anova," Memorandum N5-341, Department of Statistics, Harvard University, December 1976.

[14] Cook, R.D., "Detection of Influential Observations in Linear Regression," Technometrics, 19 (1977), pp. 15-18.

[15] Mallows, C.L., "On Some Topics in Robustness," Paper delivered at the Eastern Regional IMS meeting, University of Rochester, 1973.

Appendix 1. BASIC DIFFERENCE FORMULAS

The fundamental difference formulas are known as the Sherman-Morrison-Woodbury Theorem [19, p. 29].

$$(X_{(i)}^T X_{(i)})^{-1} = (X^T X)^{-1} + \frac{(X^T X)^{-1} x_i^T x_i (X^T X)^{-1}}{1-h_i} \tag{A1.1}$$

$$(X^T X)^{-1} = (X_{(i)}^T X_{(i)})^{-1} - \frac{(X_{(i)}^T X_{(i)})^{-1} x_i^T x_i (X_{(i)}^T X_{(i)})^{-1}}{1-x_i (X_{(i)}^T X_{(i)})^{-1} x_i^T} . \tag{A1.2}$$

From this comes

$$\hat{\beta}-\hat{\beta}_{(i)} = \frac{(X^T X)^{-1} x_i^T r_i}{1-h_i} \tag{A1.3}$$

and since

$$(n-p-1)\, s_{(i)}^2 = \sum_{\substack{t=1 \\ t\neq i}}^{n} (y_t - x_t \hat{\beta}_{(i)})^2$$

we get

$$(n-p-1)\, s_{(i)}^2 = \sum_{t=1}^{n} \left(r_t + \frac{h_{ti} r_i}{1-h_i}\right)^2 - \frac{r_i^2}{(1-h_i)^2}$$

$$= (n-p)s^2 + \frac{2r_i}{1-h_i} \sum_{t=1}^{n} r_t h_{ti} + \frac{r_i^2}{(1-h_i)^2} \sum_{t=1}^{n} h_{ti}^2 - \frac{r_i^2}{(1-h_i)^2}$$

$$= (n-p)\, s^2 - \frac{r_i^2}{1-h_i}$$

by using the fact that H annihilates the vector of residuals.

Finally we obtain

$$(n-p)\, s^2 (X^T X)^{-1} - (n-p-1)\, s^2_{(i)}\, (X^T_{(i)} X_{(i)})^{-1}$$

$$= \frac{r_i^2}{1-h_i} (X^T_{(i)} X_{(i)})^{-1} - (n-p)s^2 \frac{(X^T X)^{-1} x_i^T x_i (X^T X)^{-1}}{1-h_i} . \quad \text{(A1.4)}$$

Appendix 2. DIFFERENTIATION FORMULAS

Let

$$W = \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & w_i & \\ & & & \ddots \\ & & & & 1 \end{bmatrix} \tag{A2.1}$$

and

$$\hat{\beta}_{w_i} = (X^T W X)^{-1} X^T W Y. \tag{A2.2}$$

From (A1.1) we obtain

$$(X^T W X)^{-1} = (X^T X)^{-1} + \frac{(1-w_i)(X^T X)^{-1} x_i^T x_i (X^T X)^{-1}}{1-(1-w_i)h_i} \tag{A2.3}$$

and then

$$\frac{\partial}{\partial w_i} (X^T W X)^{-1} = \frac{-(X^T X)^{-1} x_i^T x_i (X^T X)^{-1}}{(1-(1-w_i)h_i)^2} . \tag{A2.4}$$

Some algebraic manipulation using (A2.2) and (A2.3) gives

$$\hat{\beta}_{w_i} = \hat{\beta} - (X^T X)^{-1} x_i^T r_i \frac{(1-w_i)}{1-(1-w_i)h_i} \tag{A2.5}$$

where $\hat{\beta}$ and r_i are the least-squares estimates obtained when $w_i=1$. Thus

$$\frac{\partial \beta_{w_i}}{\partial w_i} = (X^T X)^{-1} \frac{x_i^T r_i}{(1-(1-w_i)h_i)^2} \tag{A2.6}$$

or equivalently (again using A2.3)

$$= (X^T W X)^{-1} x_i^T (y_i - x_i \hat{\beta}_{w_i}). \tag{A2.7}$$

It is also useful to look at the squared residual error

$$SSR_{w_i} = \sum_{t=1}^{n} w_t (y_t - x_t \hat{\beta}_{w_i})^2 . \tag{A2.8}$$

Using (A2.7) we have

$$\frac{\partial SSR_{w_i}}{\partial w_i} = -2 \sum_{t=1}^{n} w_t (y_t - x_t \hat{\beta}_{w_i}) x_t (X^T W X)^{-1} x_i^T (y_i - x_i \hat{\beta}_{w_i})$$

$$+ (y_i - x_i \hat{\beta}_{w_i})^2$$

$$= -2 \frac{(y_i - x_i \hat{\beta}_{w_i})}{\sqrt{w_i}} \sum_{t=1}^{n} \sqrt{w_t}(y_t - x_t \hat{\beta}_{w_i}) \sqrt{w_t} x_t (X^T W X)^{-1} x_i \sqrt{w_i}$$

$$+ (y_i - x_i \hat{\beta}_{w_i})^2. \tag{A2.9}$$

For the data $\sqrt{W}\,Y$ and $\sqrt{W}\,X$

$$H_w = \sqrt{W}\, X (X^T W X)^{-1} X^T \sqrt{W} \text{ and}$$

$$H_w R_w = 0.$$

This implies that the sum in (A2.9) is zero so that

$$\frac{\partial SSR_{w_i}}{\partial w_i} = (y_i - x_i \hat{\beta}_{w_i})^2 = \frac{r_i^2}{(1-(1-w_i)h_i)^2} \tag{A2.10}$$

because of (A2.3).

Putting (A2.4) and (A2.10) together gives

$$\frac{\partial}{\partial w_i}\left[SSR_{w_i}\,(X^TWX)^{-1}\right] =$$

$$\frac{r_i^2}{(1-(1-w_i)h_i)^2}\,(X^TWX)^{-1} - SSR_{w_i}\,\frac{(X^TX)^{-1}x_i^Tx_i(X^TX)^{-1}}{(1-(1-w_i)h_i)^2}\,. \qquad \text{(A2.11)}$$

When $w_i = 1$ this is equivalent to

$$r_i^2\,(X^TX)^{-1} - (n-p)s^2\,(X^TX)^{-1}x_i^Tx_i(X^TX)^{-1}\,. \qquad \text{(A2.12)}$$

Appendix 3. THEOREMS ON THE HAT MATRIX

In this appendix we formally show that when $h_1=1$ (we can take i=1 without loss of generality), there exists a nonsingular transformation T, such that $\hat{\alpha}_1 = (T^{-1}\hat{\beta})_1 = y_1$ and $\hat{\alpha}_2,\ldots,\hat{\alpha}_p$ do not depend on y_1. This implies that, in the transformed coordinate system, the parameter α_1 has been dedicated to observation 1.

When $h_1=1$ we have for the coordinate vector $\underset{\sim}{e}_1 = (1,0,\ldots,0)^T$

$$H\underset{\sim}{e}_1 = \underset{\sim}{e}_1$$

since $h_{1j} = 0$, $j\neq 1$. Let P be any p×p nonsingular matrix whose first column is $(X^TX)^{-1}X^T\underset{\sim}{e}_1$. Then

$$XP = \begin{bmatrix} 1 & \underset{\sim}{a} \\ \underset{\sim}{0} & A \end{bmatrix}$$

where $\underset{\sim}{a}$ is 1×(p-1) and $\underset{\sim}{0}$ is (p-1)×1. Now let

$$Q = \begin{bmatrix} 1 & -\underset{\sim}{a} \\ \underset{\sim}{0} & I \end{bmatrix}$$

with I denoting the (p-1)×(p-1) identity matrix. The transformation we seek is given by T = PQ, which is nonsingular because both P and Q have inverses. Clearly

$$XT = \begin{bmatrix} 1 & \underset{\sim}{0} \\ \underset{\sim}{0} & A \end{bmatrix} ,$$

and the least-squares estimate of the parameter $\underset{\sim}{\alpha} = T^{-1}\underset{\sim}{\beta}$ will have the first residual, $y_1-\hat{\alpha}_1$, equal to zero since $\hat{\alpha}_2,\ldots,\hat{\alpha}_p$ cannot affect this residual. This also implies that $\hat{\alpha}_2,\ldots,\hat{\alpha}_p$ will not depend on y_1.

To prove the second theorem in Section 2.3

$$\det(X_{(i)}{}^{T}X_{(i)}) = (1-h_i)\det(X^{T}X)$$

we need first to show that

$$\det(I-uv^{T}) = 1-v^{T}u$$

where u and v are column vectors. Let Q be an orthonormal matrix such that

$$Qu = ||u||e_1 \qquad \text{(A3.1)}$$

where e_1 is the first standard basis vector. Then

$$\det(I-uv^{T}) = \det Q[I-uv^{T}]\,Q^{T}$$

$$= \det[I-||u||e_1v^{T}Q^{T}] = 1 - v^{T}Q^{T}e_1\,||u||$$

which is just $1-v^{T}u$ because of (A3.1). Now

$$\det X_{(i)}{}^{T}X_{(i)} = \det[(I-x_i^{T}x_i(X^{T}X)^{-1})\,X^{T}X]$$

and letting $u = x_i^{T}$ and $v = x_i(X^{T}X)^{-1}$ completes the proof since $x_i(X^{T}X)^{-1}x_i^{T}=h_i$. (We are indebted to David Gay for simplifying our original proof.)

Appendix 4. EXHIBITS FOR SECTION 2.9

Exhibit No.	Title
1	Assignments of Row Indices to Countries
2	Data
3	Ordinary Least Squares Regression Results
4	Normal Probability Plot of Studentized Residuals
5	Studentized Residuals
6	Diagonal Elements of the Hat Matrix.
7	NBFBETAS: Square Roots of the Sum of Squares of the Scaled Differences of LS Full Data and Row Removed Coefficients (DFBETAS)
8 - 11	DFBETAS (for individual coefficients)
12	Summary of Relative Changes in Coefficient Standard Errors: NDFBVARS
13 - 16	Individual Relative Change in Coefficient Standard Errors: DFBVARS
17	Scaled Change in Fit
18	Scatter Plot of NDFBETAS versus Diagonal Elements of the Hat Matrix
19	Scatter Plot of NDFBETAS versus Studentized Residuals
20	Scatter Plot of NDFBVARS versus Diagonal Elements of the Hat Matrix
21	Scatter Plot of NDFBETAS versus NDFBVARS
22	Diagonals of Hat Matrix with Observation 49 Removed
23	NDFBETAS with Observation 49 Removed

EXHIBIT 1

POSITION	LABEL
1	AUSTRALIA
2	AUSTRIA
3	BELGIUM
4	BOLIVIA
5	BRAZIL
6	CANADA
7	CHILE
8	CHINA(TAIWAN)
9	COLOMBIA
10	COSTA RICA
11	DENMARK
12	ECUADOR
13	FINLAND
14	FRANCE
15	GERMANY F.R.
16	GREECE
17	GUATEMALA
18	HONDURAS
19	ICELAND
20	INDIA
21	IRELAND
22	ITALY
23	JAPAN
24	KOREA
25	LUXEMBOURG
26	MALTA
27	NORWAY
28	NETHERLANDS
29	NEW ZEALAND
30	NICARAGUA
31	PANAMA
32	PARAGUAY
33	PERU
34	PHILLIPINES
35	PORTUGAL
36	SOUTH AFRICA
37	SOUTH RHODESIA
38	SPAIN
39	SWEDEN
40	SWITZERLAND
41	TURKEY
42	TUNISIA
43	UNITED KINGDOM
44	UNITED STATES
45	VENEZUELA
46	ZAMBIA
47	JAMAICA
48	URUGUAY
49	LIBYA
50	MALAYSIA

EXHIBIT 2

	Y	COL 2	COL 3	COL 4
AUSTRALIA	11.43	29.35	2.87	2329.68
AUSTRIA	12.07	23.32	4.41	1507.99
BELGIUM	13.17	23.8	4.43	2108.47
BOLIVIA	5.75	41.89	1.67	189.13
BRAZIL	12.88	42.19	0.83	728.47
CANADA	8.79	31.72	2.85	2982.88
CHILE	0.6	39.74	1.34	662.86
CHINA(TAIWAN)	11.9	44.75	0.67	289.52
COLOMBIA	4.98	46.64	1.06	276.65
COSTA RICA	10.78	47.64	1.14	471.24
DENMARK	16.85	24.42	3.93	2496.53
ECUADOR	3.59	46.31	1.19	287.77
FINLAND	11.24	27.84	2.37	1681.25
FRANCE	12.64	25.06	4.7	2213.82
GERMANY F.R.	12.55	23.31	3.35	2457.12
GREECE	10.67	25.62	3.1	870.85
GUATEMALA	3.01	46.05	0.87	289.71
HONDURAS	7.7	47.32	0.58	232.44
ICELAND	1.27	34.03	3.08	1900.1
INDIA	9.	41.31	0.96	88.94
IRELAND	11.34	31.16	4.19	1139.95
ITALY	14.28	24.52	3.48	1390.
JAPAN	21.1	27.01	1.91	1257.28
KOREA	3.98	41.74	0.91	207.6
LUXEMBOURG	10.35	21.8	3.73	2449.39
MALTA	15.48	32.54	2.47	601.05
NORWAY	10.25	25.95	3.67	2231.03
NETHERLANDS	14.65	24.71	3.25	1740.7
NEW ZEALAND	10.67	32.61	3.17	1487.52
NICARAGUA	7.3	45.04	1.21	325.54
PANAMA	4.44	43.56	1.2	568.56
PARAGUAY	2.02	41.18	1.05	220.56
PERU	12.7	44.19	1.28	400.06
PHILLIPINES	12.78	46.26	1.12	152.01
PORTUGAL	12.49	28.96	2.85	579.51
SOUTH AFRICA	11.14	31.94	2.28	651.11
SOUTH RHODESIA	13.3	31.92	1.52	250.96
SPAIN	11.77	27.74	2.87	768.79
SWEDEN	6.86	21.44	4.54	3299.49
SWITZERLAND	14.13	23.49	3.73	2630.96
TURKEY	5.13	43.42	1.08	389.66
TUNISIA	2.81	46.12	1.21	249.87
UNITED KINGDOM	7.81	23.27	4.46	1813.93
UNITED STATES	7.56	29.81	3.43	4001.89
VENEZUELA	9.22	46.4	0.9	813.39
ZAMBIA	18.56	45.25	0.56	138.33
JAMAICA	7.72	41.12	1.73	380.47
URUGUAY	9.24	28.13	2.72	766.54
LIBYA	8.89	43.69	2.07	123.5
MALAYSIA	4.71	47.2	0.66	242.69

EXHIBIT 1

POSITION	LABEL
1	AUSTRALIA
2	AUSTRIA
3	BELGIUM
4	BOLIVIA
5	BRAZIL
6	CANADA
7	CHILE
8	CHINA(TAIWAN)
9	COLOMBIA
10	COSTA RICA
11	DENMARK
12	ECUADOR
13	FINLAND
14	FRANCE
15	GERMANY F.R.
16	GREECE
17	GUATEMALA
18	HONDURAS
19	ICELAND
20	INDIA
21	IRELAND
22	ITALY
23	JAPAN
24	KOREA
25	LUXEMBOURG
26	MALTA
27	NORWAY
28	NETHERLANDS
29	NEW ZEALAND
30	NICARAGUA
31	PANAMA
32	PARAGUAY
33	PERU
34	PHILLIPINES
35	PORTUGAL
36	SOUTH AFRICA
37	SOUTH RHODESIA
38	SPAIN
39	SWEDEN
40	SWITZERLAND
41	TURKEY
42	TUNISIA
43	UNITED KINGDOM
44	UNITED STATES
45	VENEZUELA
46	ZAMBIA
47	JAMAICA
48	URUGUAY
49	LIBYA
50	MALAYSIA

EXHIBIT 2 CONTINUED

	COL 5
AUSTRALIA	2.87
AUSTRIA	3.93
BELGIUM	3.82
BOLIVIA	0.22
BRAZIL	4.56
CANADA	2.43
CHILE	2.67
CHINA(TAIWAN)	6.51
COLOMBIA	3.08
COSTA RICA	2.8
DENMARK	3.99
ECUADOR	2.19
FINLAND	4.32
FRANCE	4.52
GERMANY F.R.	3.44
GREECE	6.28
GUATEMALA	1.48
HONDURAS	3.19
ICELAND	1.12
INDIA	1.54
IRELAND	2.99
ITALY	3.54
JAPAN	8.21
KOREA	5.81
LUXEMBOURG	1.57
MALTA	8.12
NORWAY	3.62
NETHERLANDS	7.66
NEW ZEALAND	1.76
NICARAGUA	2.48
PANAMA	3.61
PARAGUAY	1.03
PERU	0.67
PHILLIPINES	2.
PORTUGAL	7.48
SOUTH AFRICA	2.19
SOUTH RHODESIA	2.
SPAIN	4.35
SWEDEN	3.01
SWITZERLAND	2.7
TURKEY	2.96
TUNISIA	1.13
UNITED KINGDOM	2.01
UNITED STATES	2.45
VENEZUELA	0.53
ZAMBIA	5.14
JAMAICA	10.23
URUGUAY	1.88
LIBYA	16.71
MALAYSIA	5.08

EXHIBIT 3

NOB=50 NOVAR=5 CONDITION OF SCALED X=34.8683

SSR=650.713 SER=3.80267 RSQ=.338456 DW(0)=1.9234

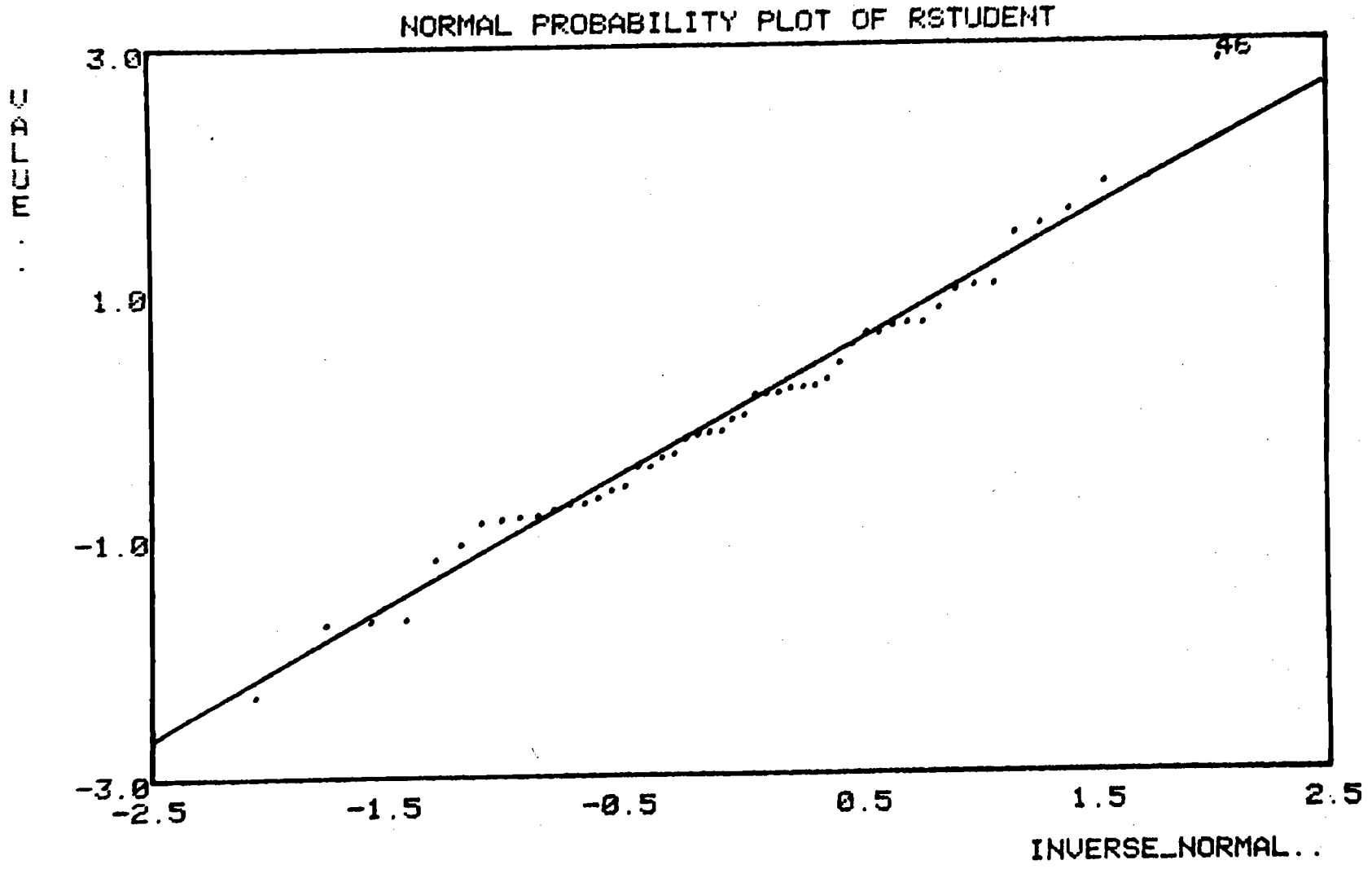
F(4,45)=5.75568 TRACE OF XTXINV=3.82582

	NAME	EST COEF	STD ERR	T-STAT
COEF.1	INTERCEPT	28.5661	7.35452	3.88415
COEF.2	POP15	-0.461193	0.144642	-3.18851
COEF.3	POP75	-1.6915	1.0836	-1.561
COEF.4	INC	-0.000337	0.000931	-0.361829
COEF.5	INGRO	0.409694	0.196197	2.08818

EXHIBIT 2 CONTINUED

	COL 5
AUSTRALIA	2.87
AUSTRIA	3.93
BELGIUM	3.82
BOLIVIA	0.22
BRAZIL	4.56
CANADA	2.43
CHILE	2.67
CHINA(TAIWAN)	6.51
COLOMBIA	3.08
COSTA RICA	2.8
DENMARK	3.99
ECUADOR	2.19
FINLAND	4.32
FRANCE	4.52
GERMANY F.R.	3.44
GREECE	6.28
GUATEMALA	1.48
HONDURAS	3.19
ICELAND	1.12
INDIA	1.54
IRELAND	2.99
ITALY	3.54
JAPAN	8.21
KOREA	5.81
LUXEMBOURG	1.57
MALTA	8.12
NORWAY	3.62
NETHERLANDS	7.66
NEW ZEALAND	1.76
NICARAGUA	2.48
PANAMA	3.61
PARAGUAY	1.03
PERU	0.67
PHILLIPINES	2.
PORTUGAL	7.48
SOUTH AFRICA	2.19
SOUTH RHODESIA	2.
SPAIN	4.35
SWEDEN	3.01
SWITZERLAND	2.7
TURKEY	2.96
TUNISIA	1.13
UNITED KINGDOM	2.01
UNITED STATES	2.45
VENEZUELA	0.53
ZAMBIA	5.14
JAMAICA	10.23
URUGUAY	1.88
LIBYA	16.71
MALAYSIA	5.08

EXHIBIT 4



ROBUST EQUATION IS: Y=1.05806 X-0.018218

EXHIBIT 5

INDEX PLOT OF RSTUDENT

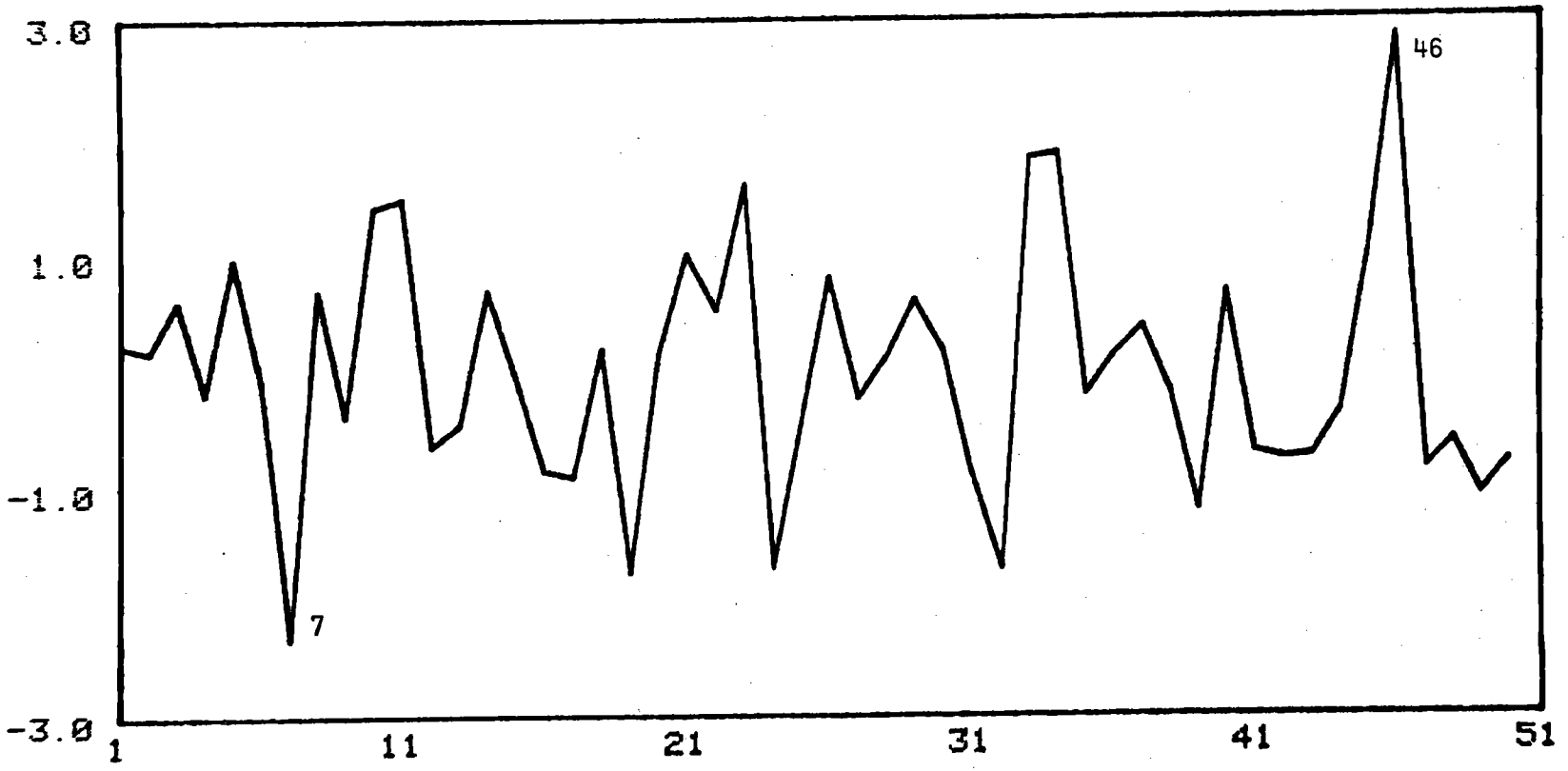
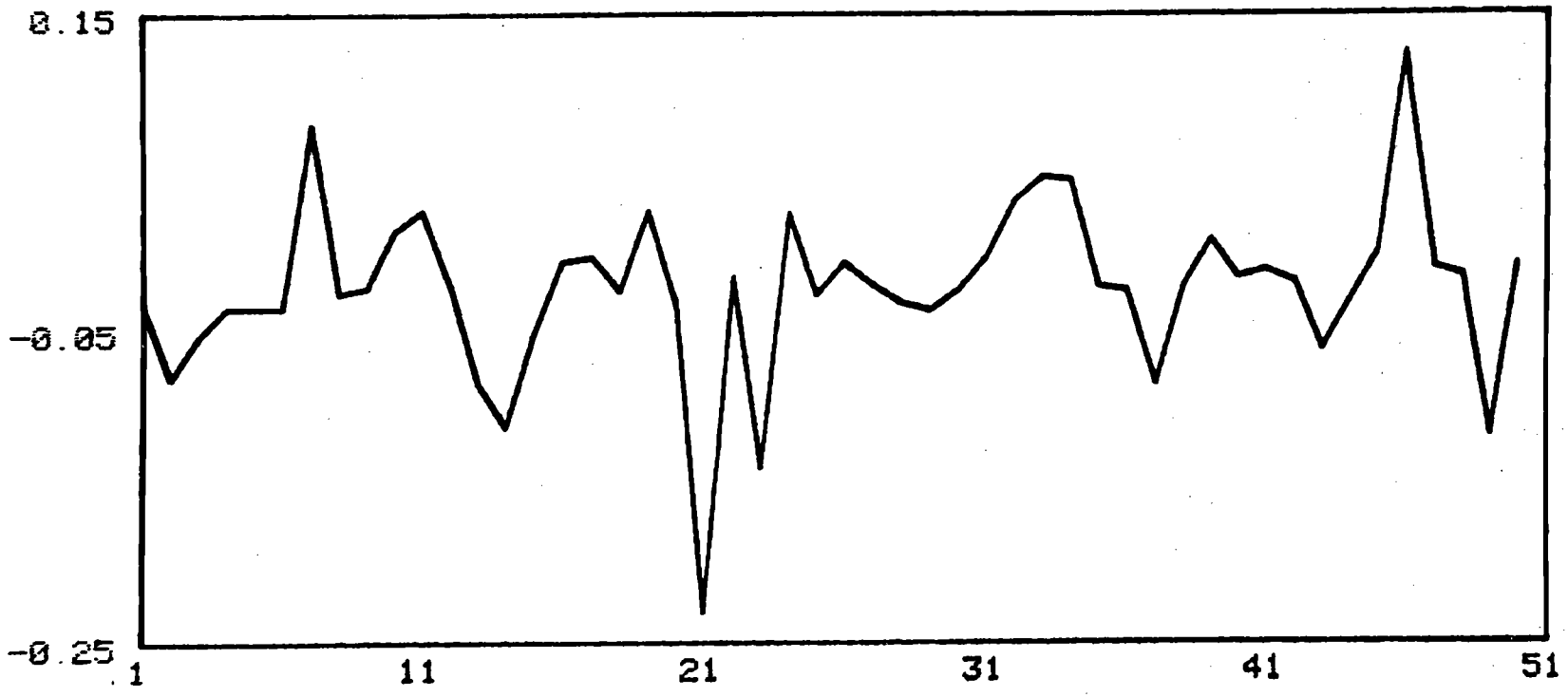


EXHIBIT 14

INDEX PLOT OF SELECTED COLUMNS OF DFBVARS



TIME BOUNDS: 1 TO 50

DATA NAMES: C3(POP75)

EXHIBIT 7

INDEX PLOT OF NOFBETAS

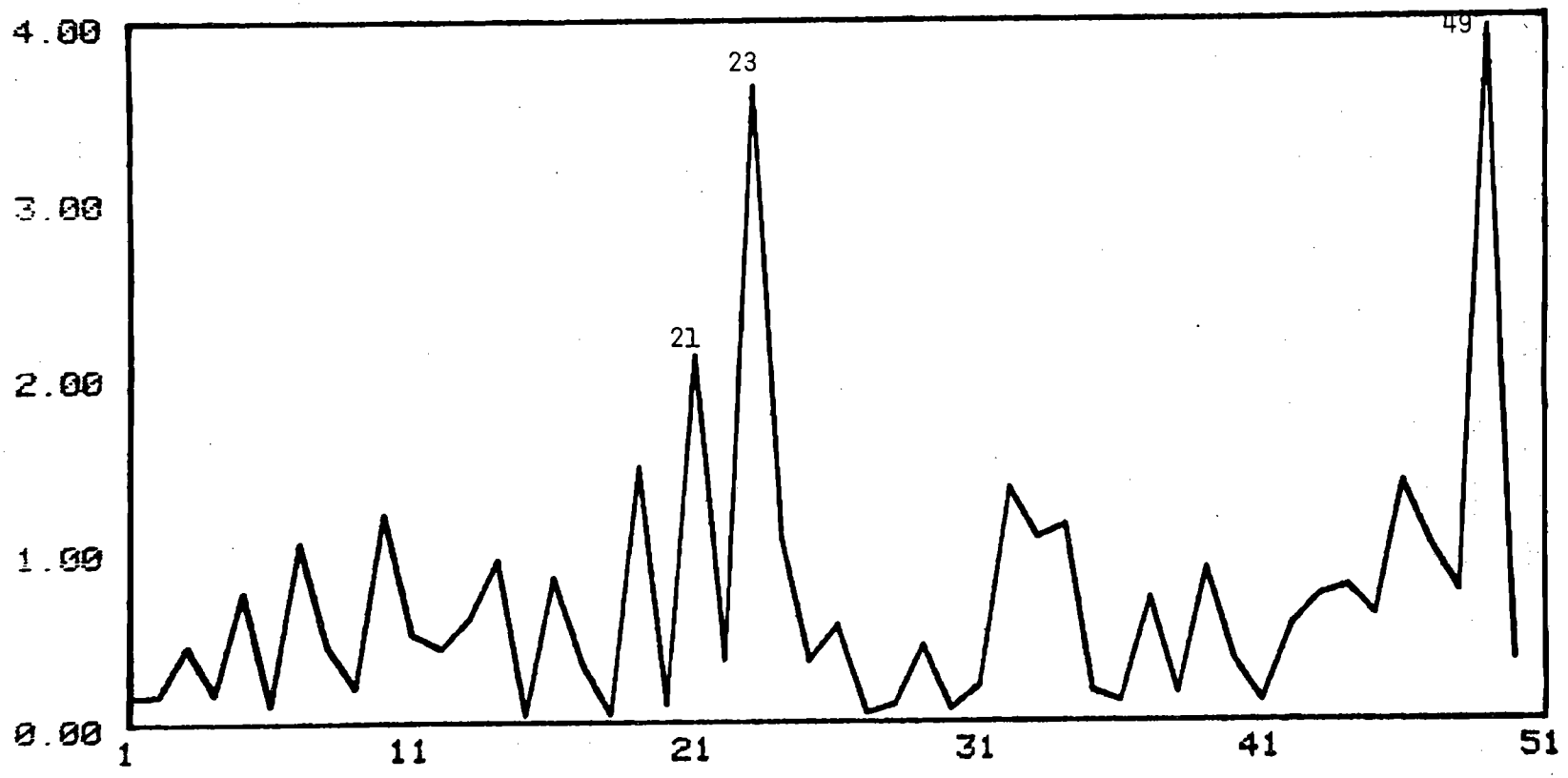
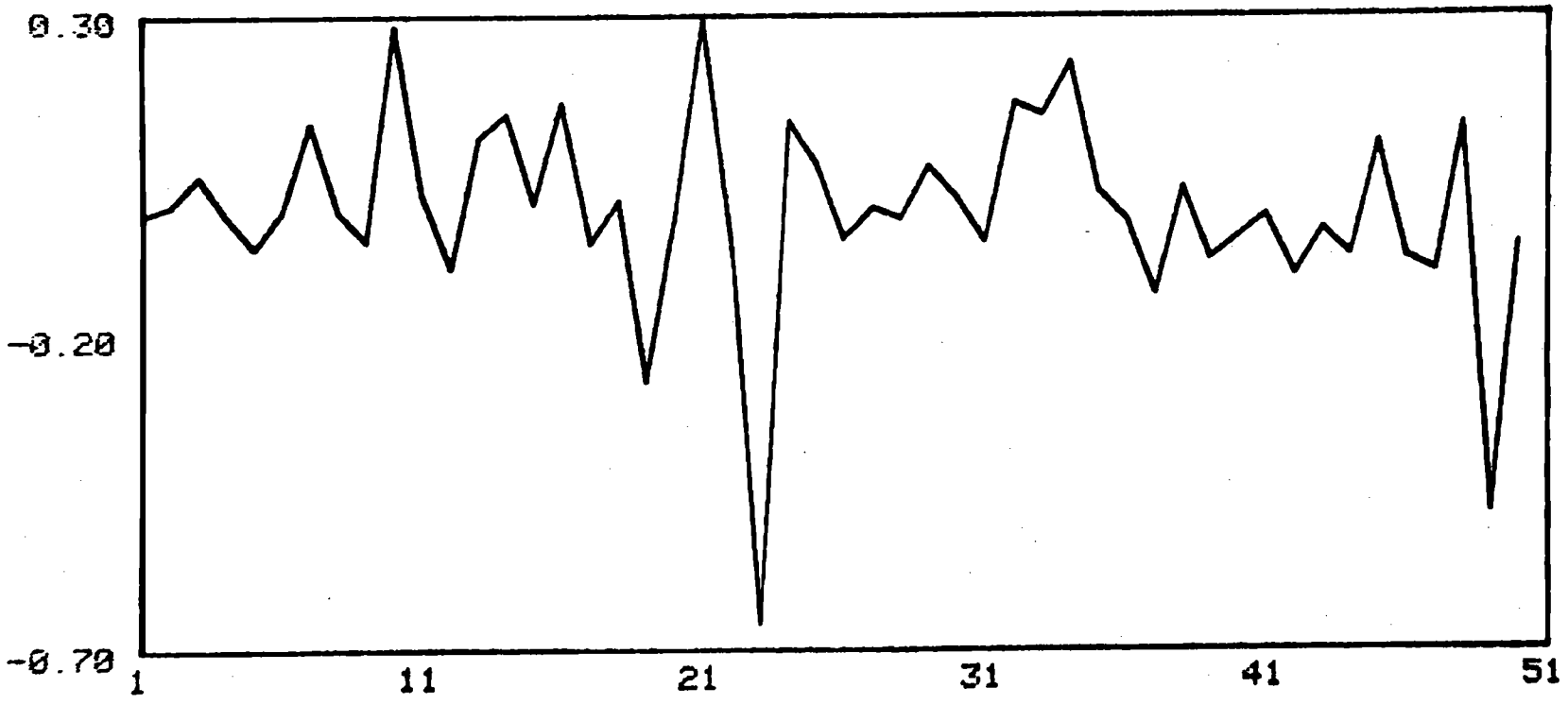


EXHIBIT 8

INDEX PLOT OF SELECTED COLUMNS OF DFBETAS

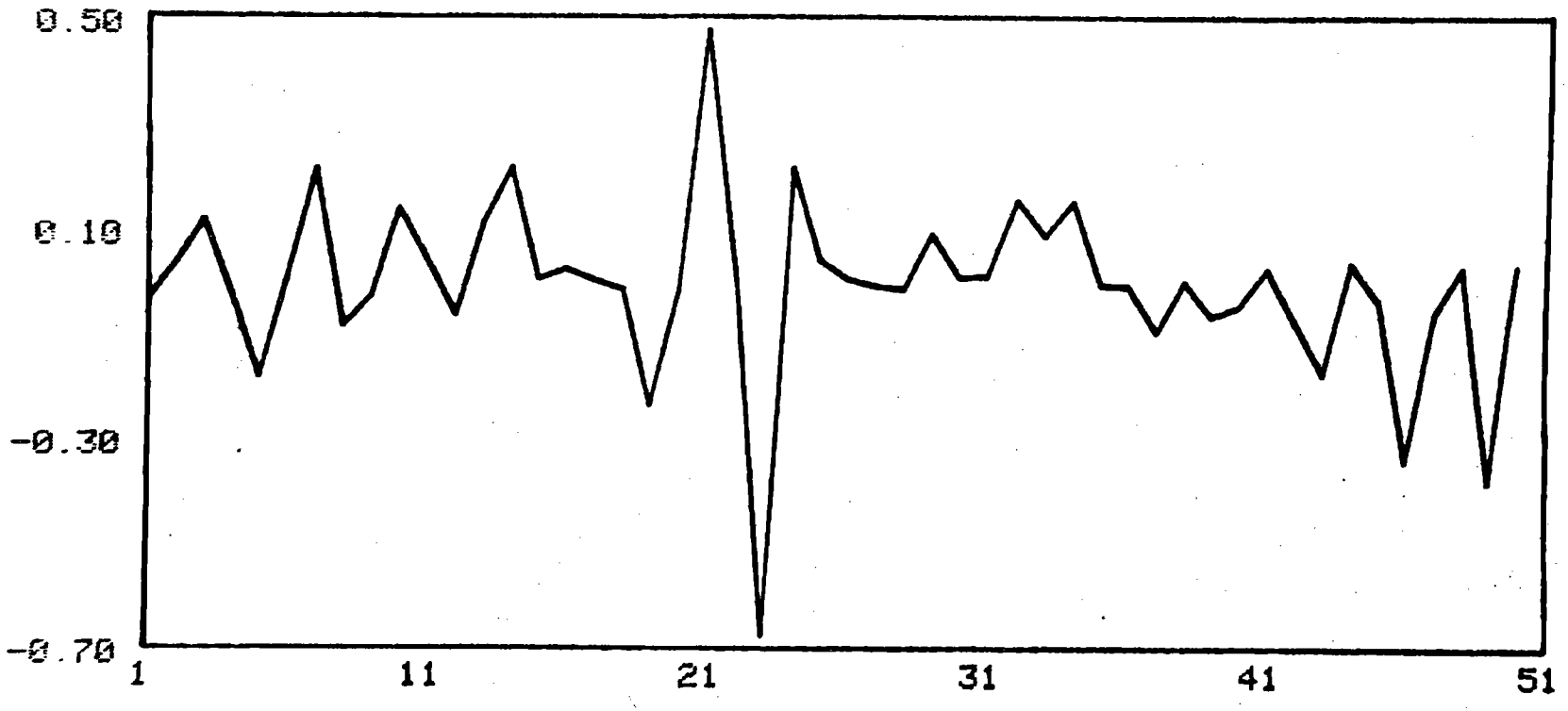


TIME BOUNDS: 1 TO 50

DATA NAMES: C2(POP15)

EXHIBIT 9

INDEX PLOT OF SELECTED COLUMNS OF DFBETAS

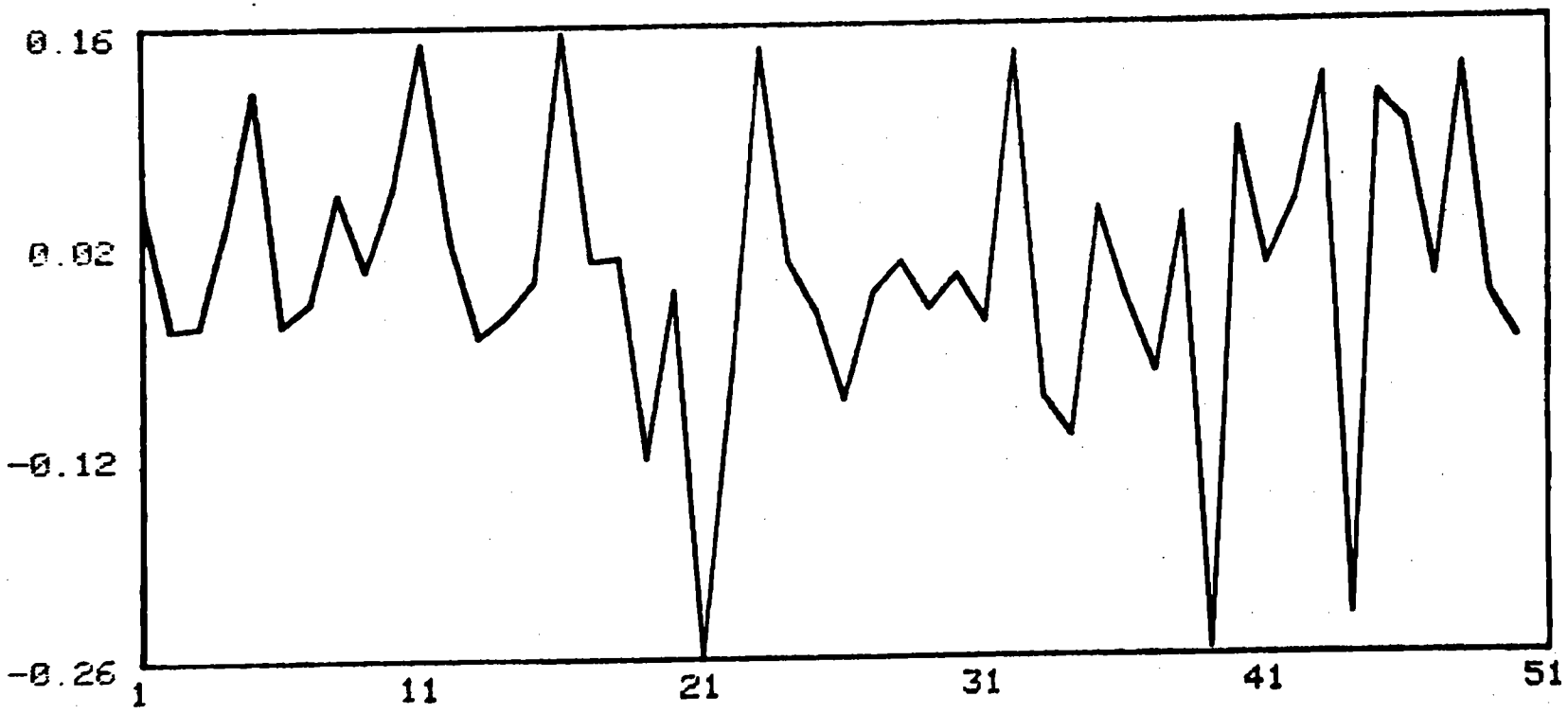


TIME BOUNDS: 1 TO 50

DATA NAMES: C3 (POP75)

EXHIBIT 10

INDEX PLOT OF SELECTED COLUMNS OF DFBETAS

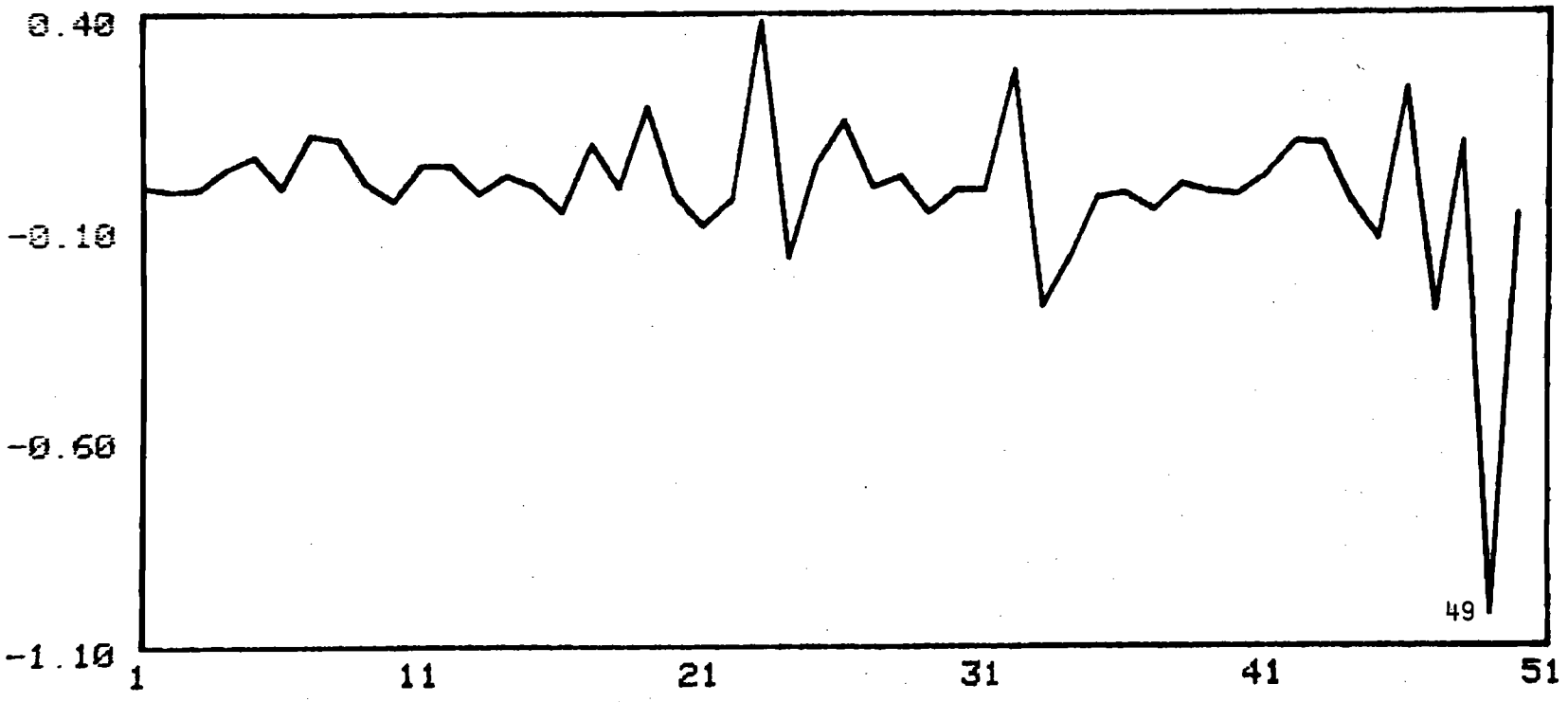


TIME BOUNDS: 1 TO 50

DATA NAMES: C4(INC)

EXHIBIT 11

INDEX PLOT OF SELECTED COLUMNS OF DFBETAS



TIME BOUNDS: 1 TO 50

DATA NAMES: C5 (INGRO)

EXHIBIT 12

INDEX PLOT OF NOFBVARS

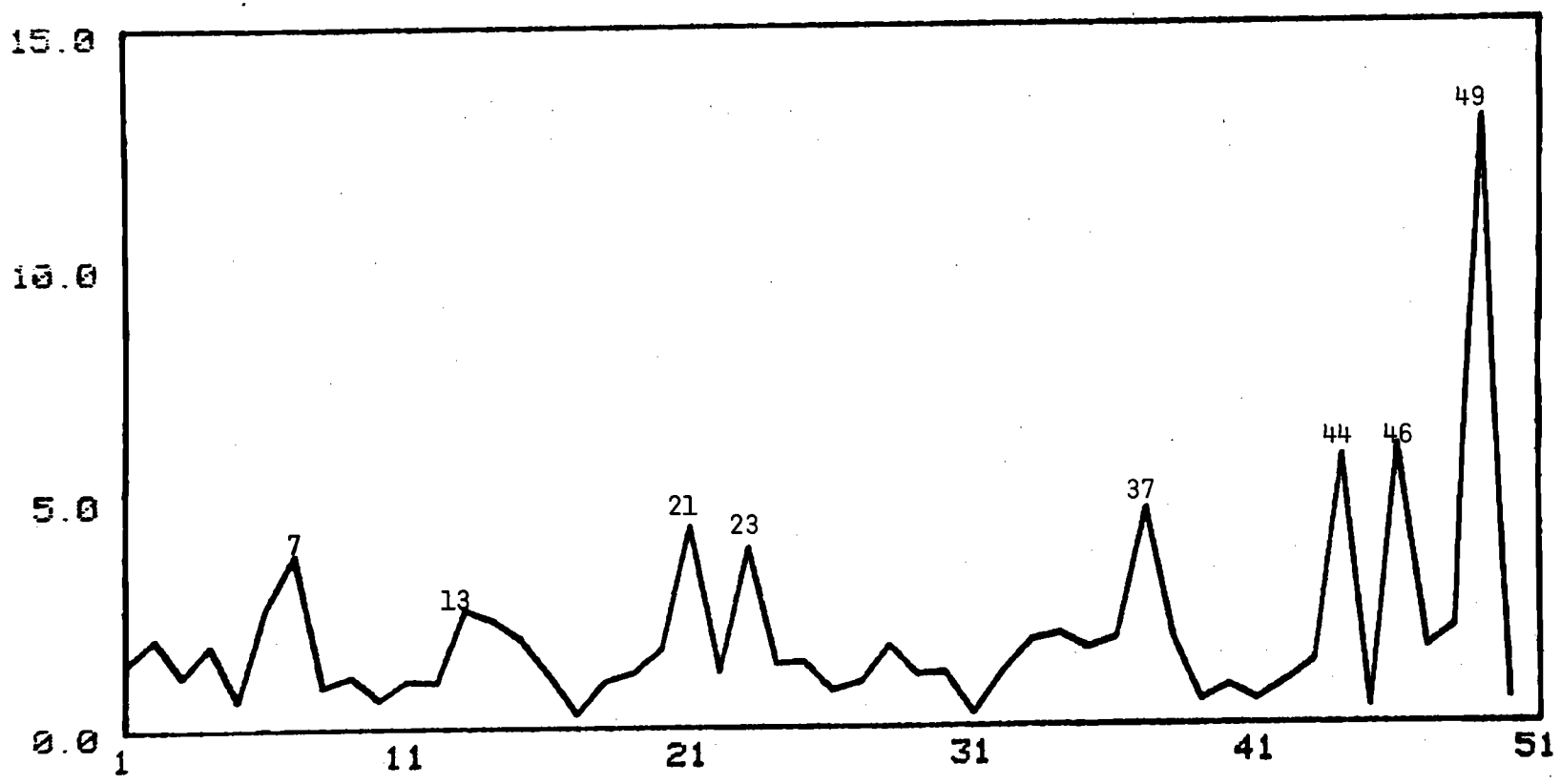
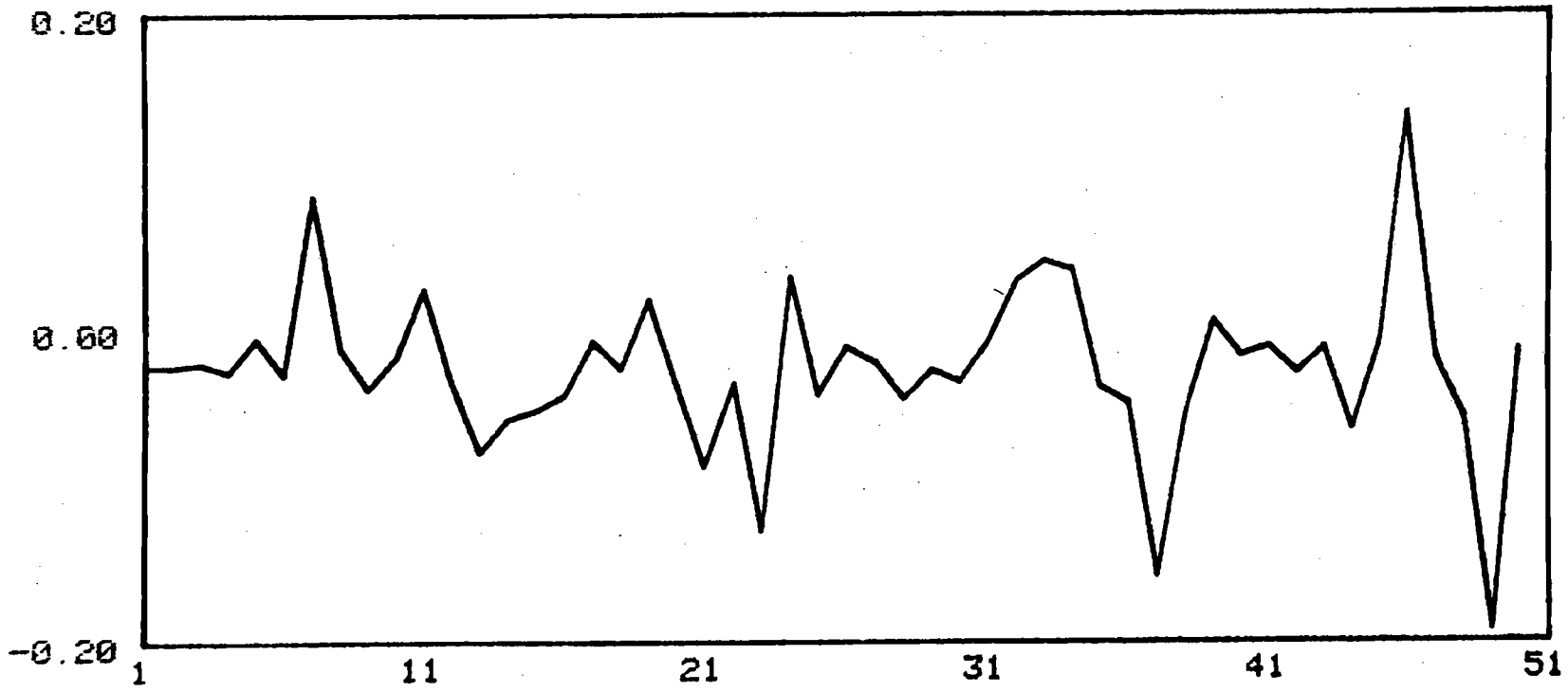


EXHIBIT 13

INDEX PLOT OF SELECTED COLUMNS OF OFBVARS

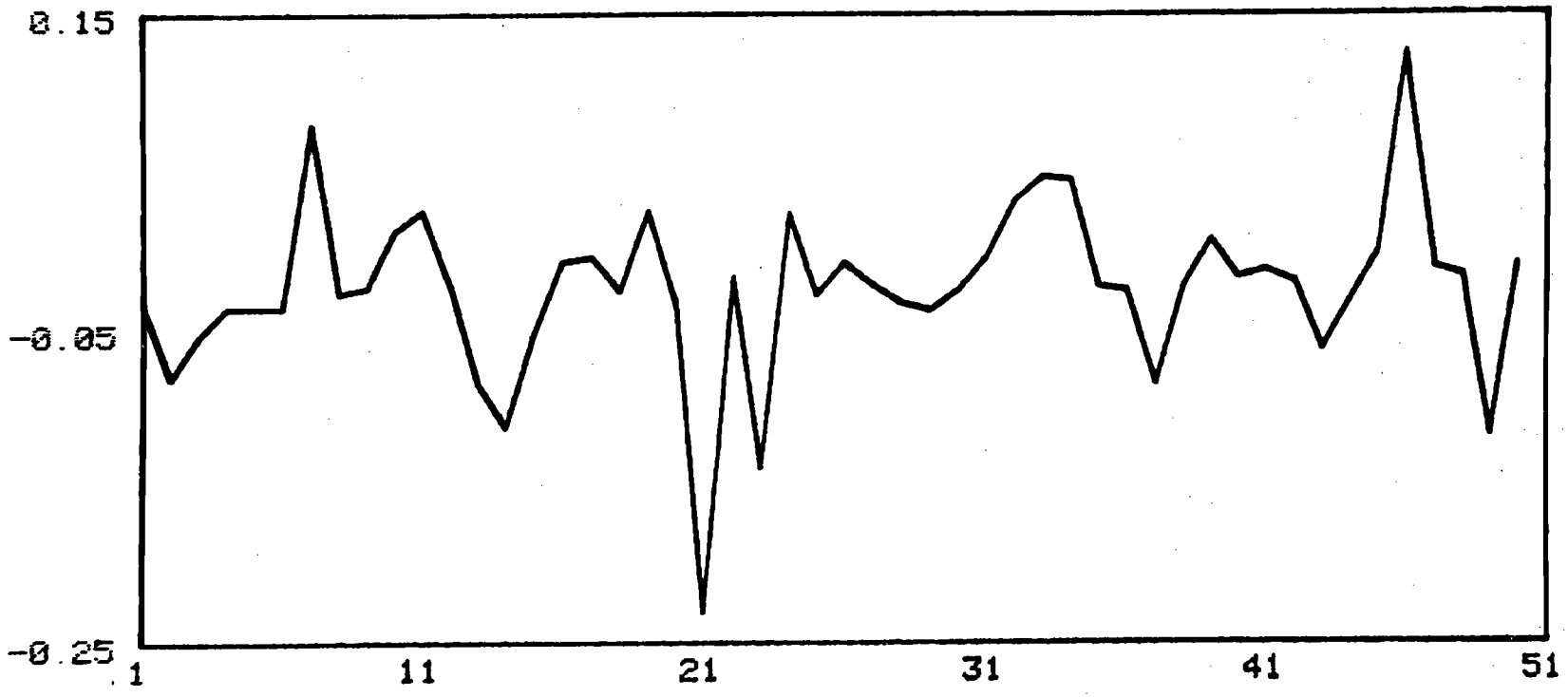


TIME BOUNDS: 1 TO 50

DATA NAMES: C2(POP15)

EXHIBIT 14

INDEX PLOT OF SELECTED COLUMNS OF DFBVARS

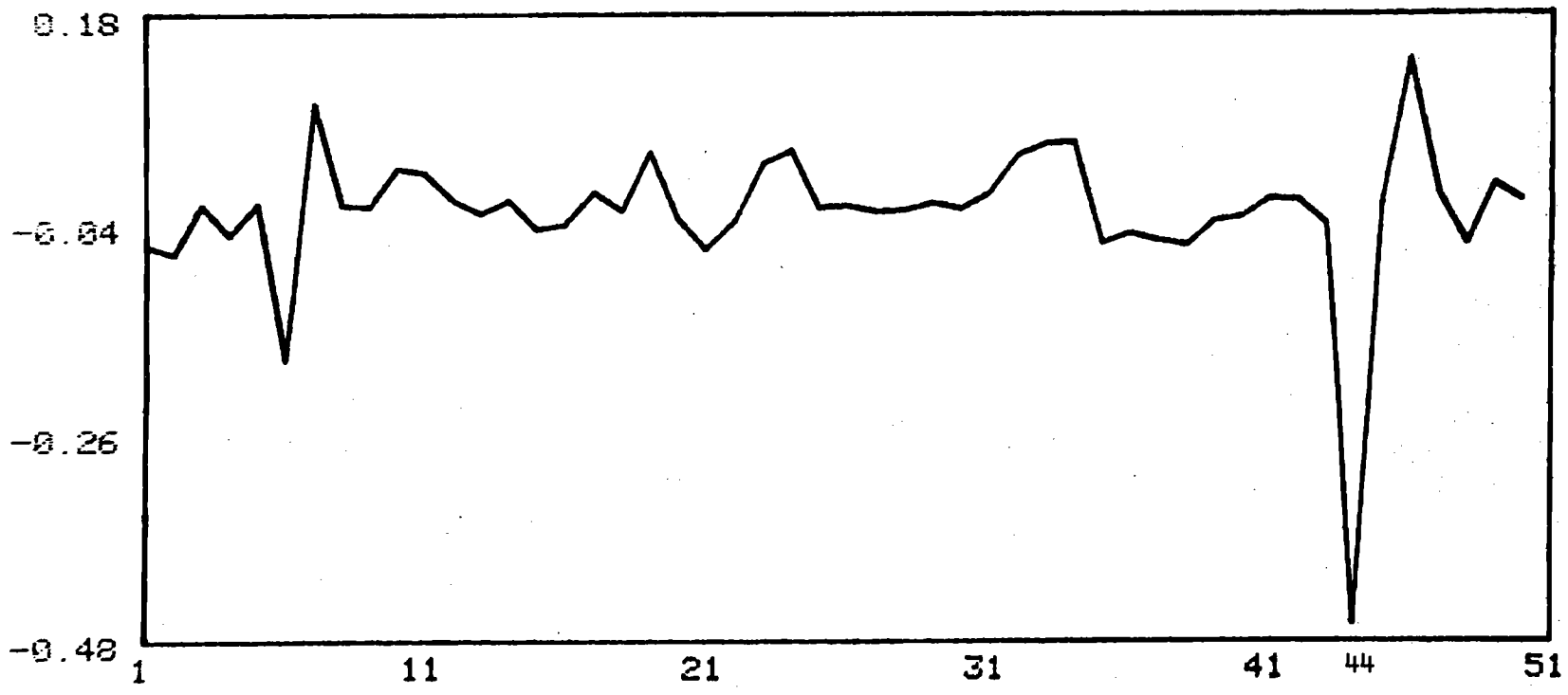


TIME BOUNDS: 1 TO 50

DATA NAMES: C3(POP75)

EXHIBIT 15

INDEX PLOT OF SELECTED COLUMNS OF DFBVARS

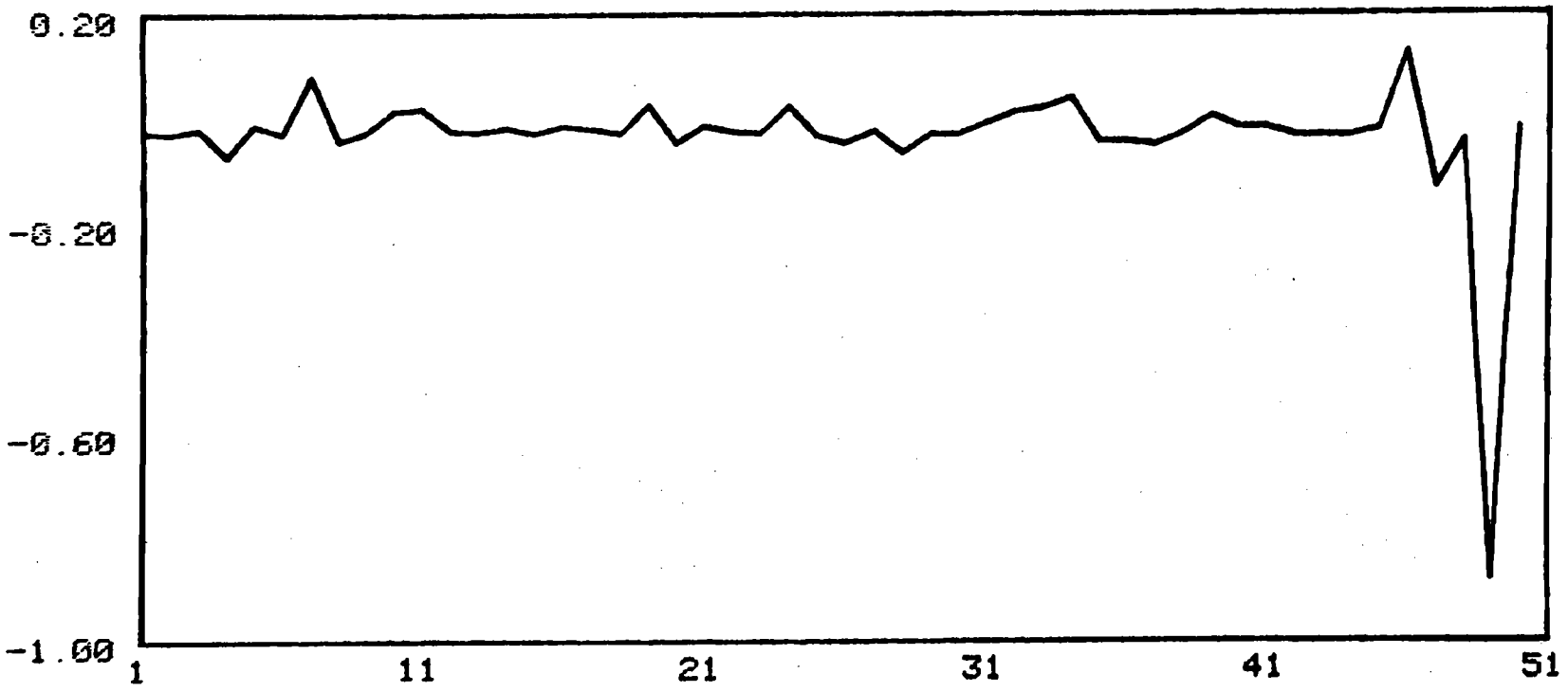


TIME BOUNDS: 1 TO 50

DATA NAMES: C4 (INC)

EXHIBIT 16

INDEX PLOT OF SELECTED COLUMNS OF DFBVARS



TIME BOUNDS: 1 TO 50

DATA NAMES: C5(INGRO)

EXHIBIT 17

INDEX PLOT OF DFFITS

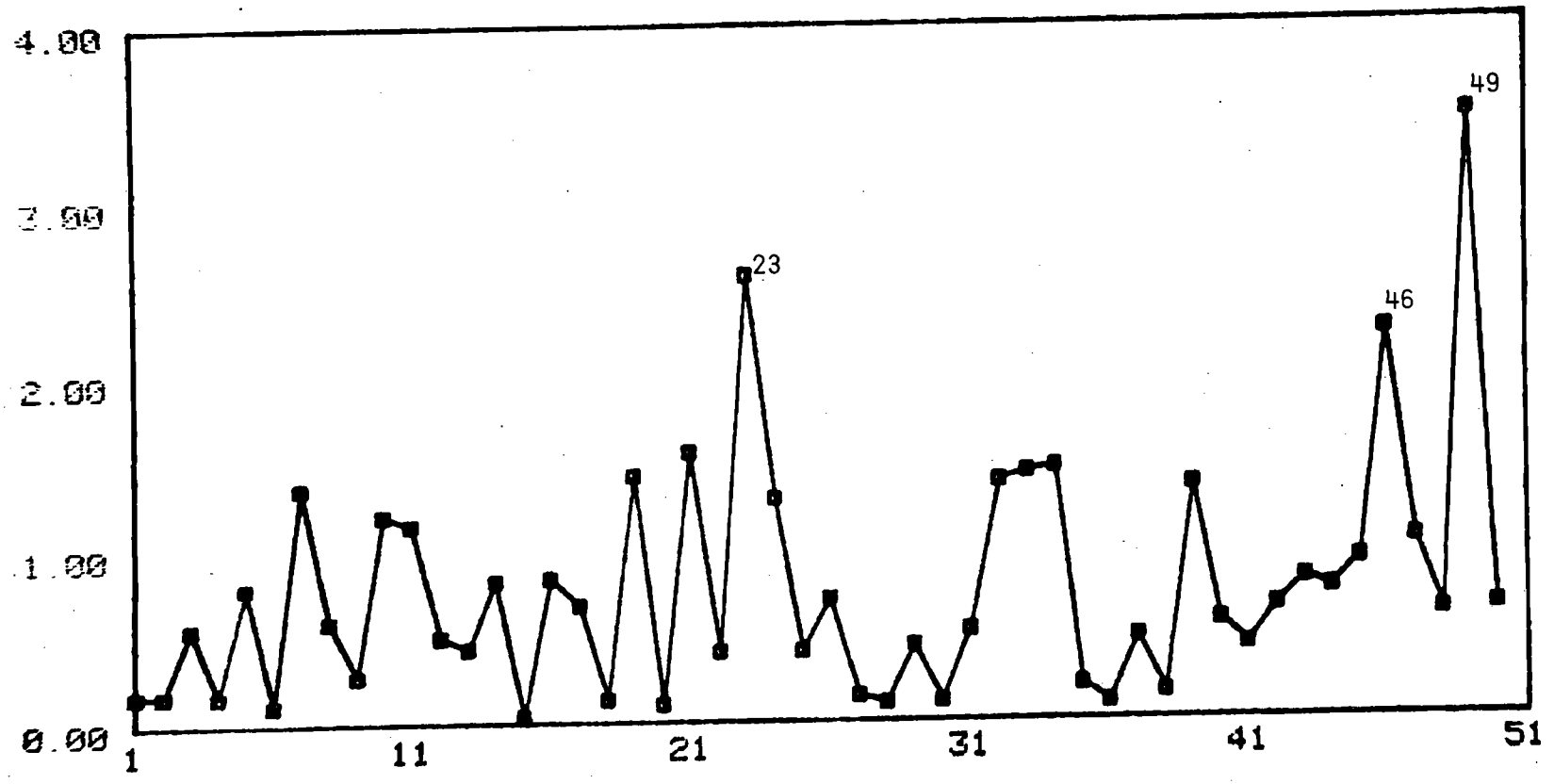


EXHIBIT 18

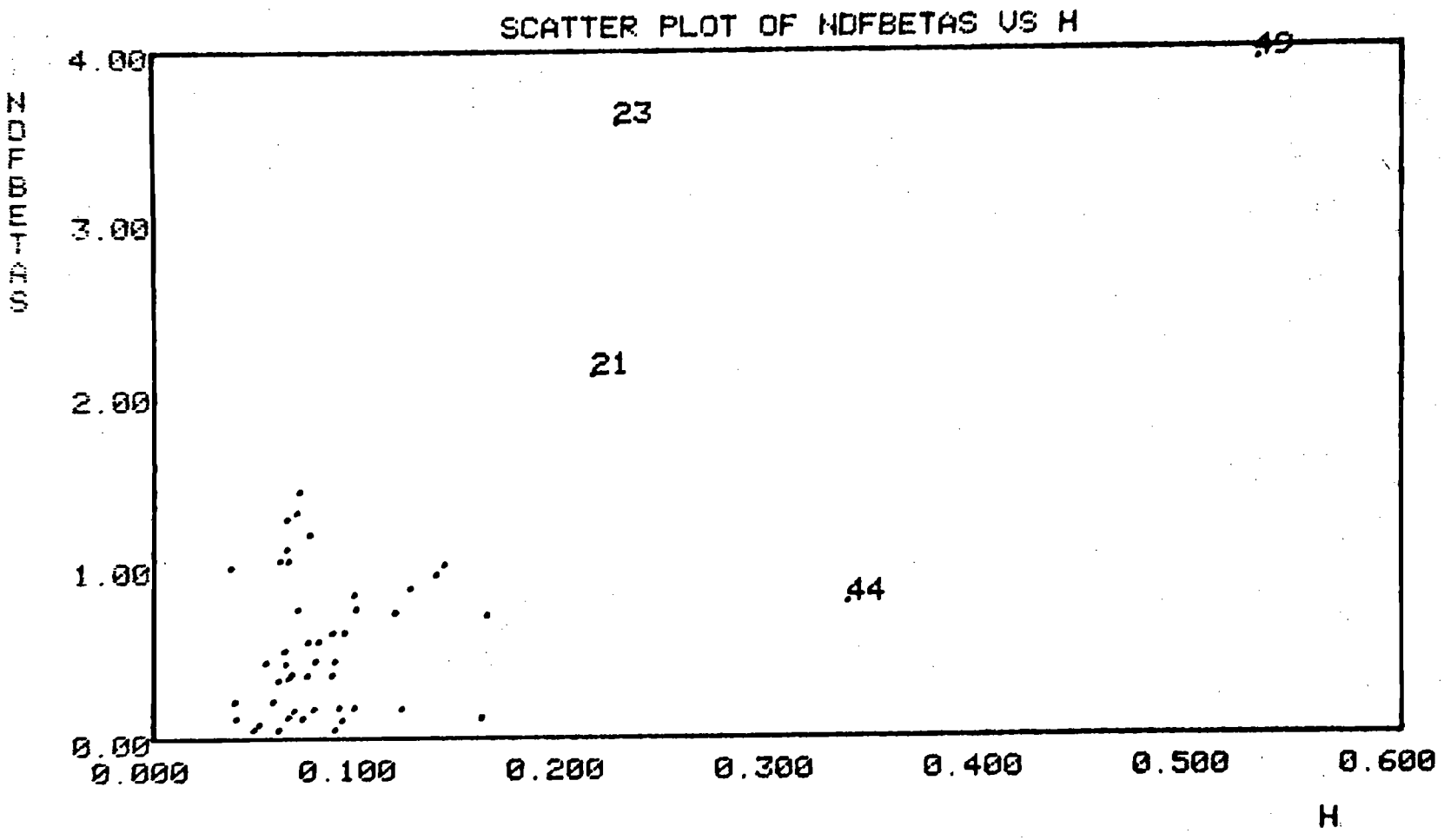

SCATTER PLOT OF NDFBETAS VS H
NDFBETAS
4.00
3.00
2.00
1.00
0.00
0.000
0.100
0.200
0.300
0.400
0.500
0.600
H
19
23
21
44

EXHIBIT 19

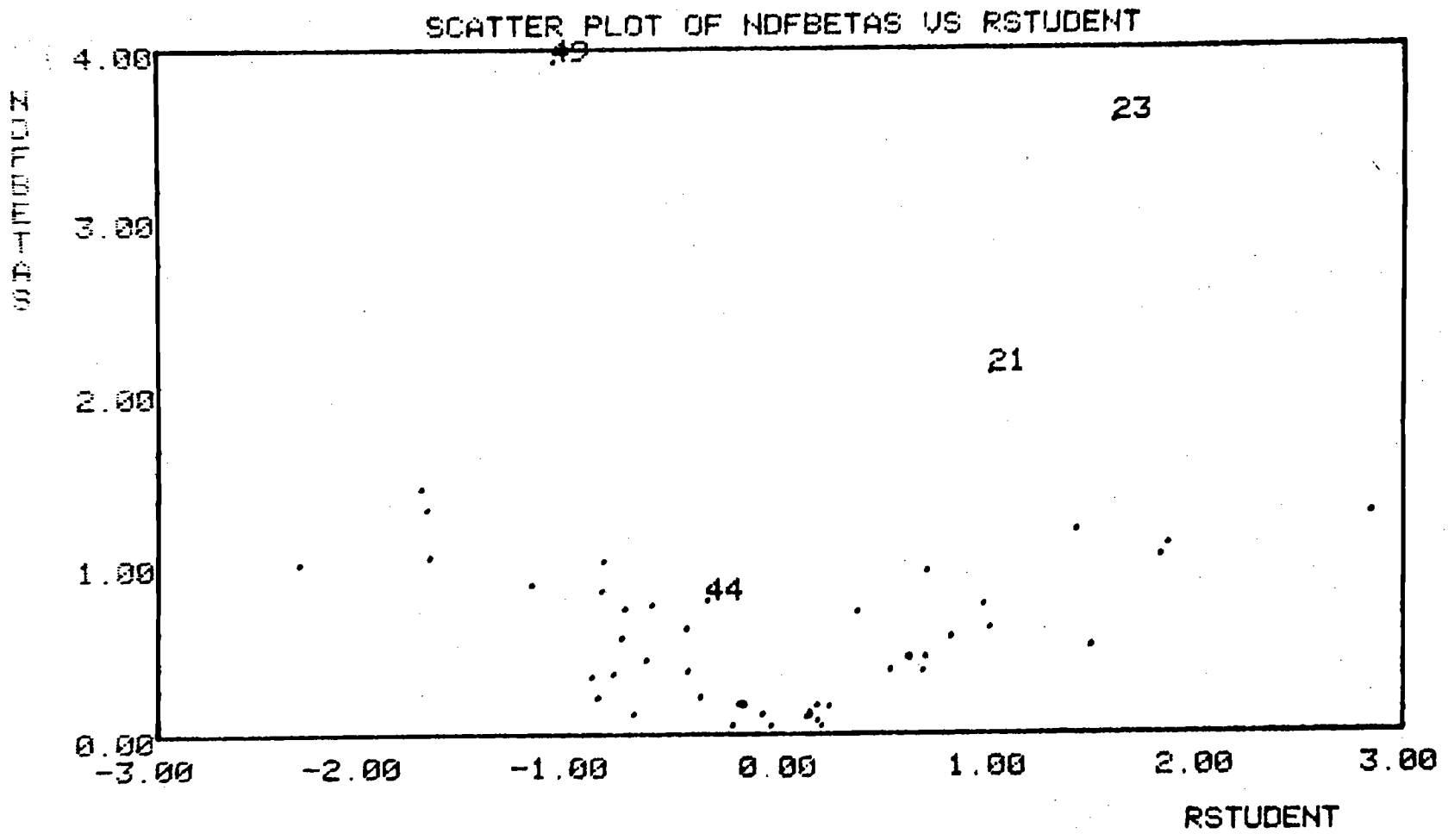

SCATTER PLOT OF NDFBETAS VS RSTUDENT
NDFBETAS
4.00
3.00
2.00
1.00
0.00
49
23
21
44
-3.00
-2.00
-1.00
0.00
1.00
2.00
3.00
RSTUDENT

EXHIBIT 20

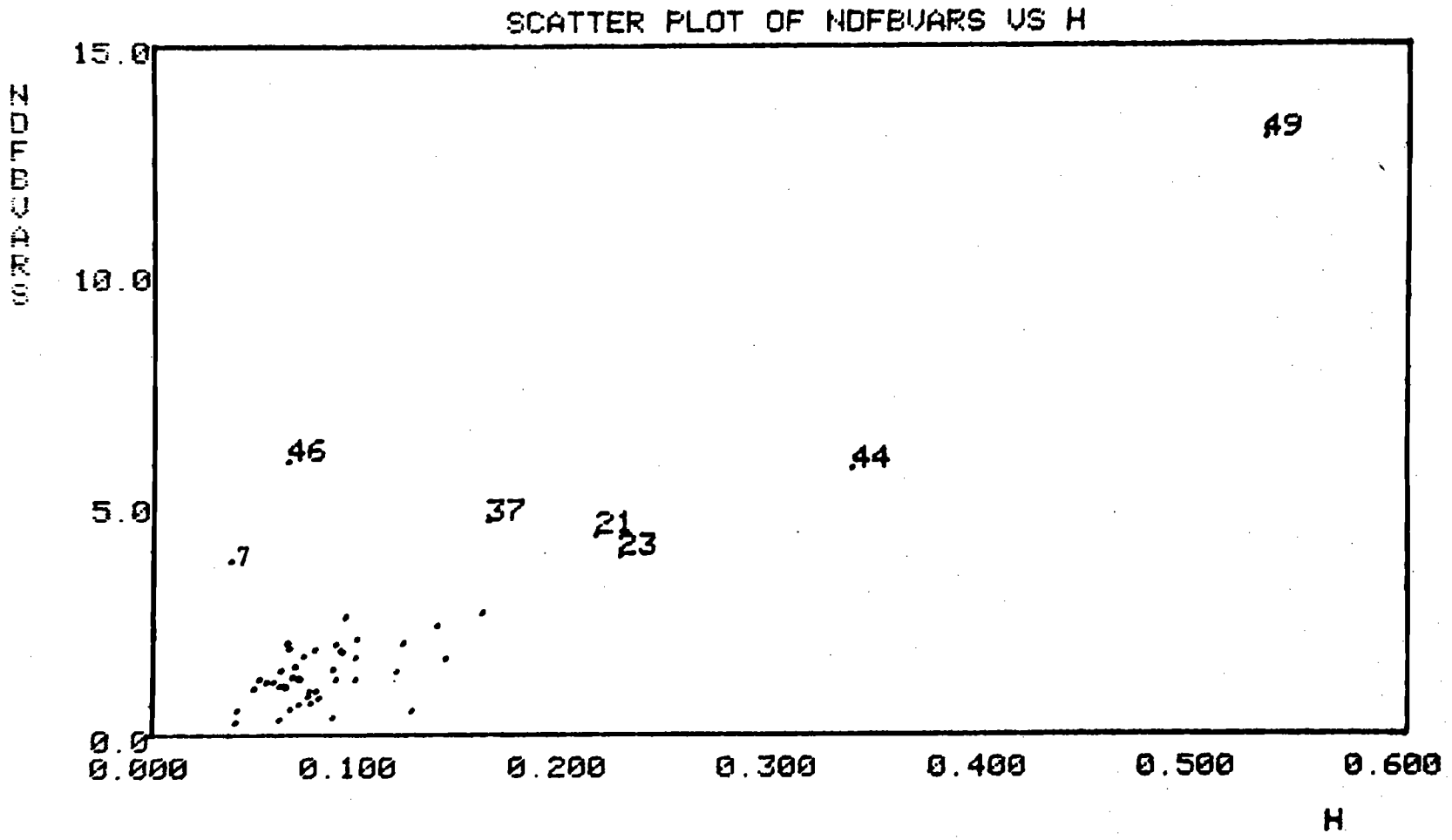

SCATTER PLOT OF NDFBUARS VS H
NDFBUARS
15.0
10.0
5.0
0.0
0.000
0.100
0.200
0.300
0.400
0.500
0.600
H
49
46
44
37
21
23
7

EXHIBIT 21

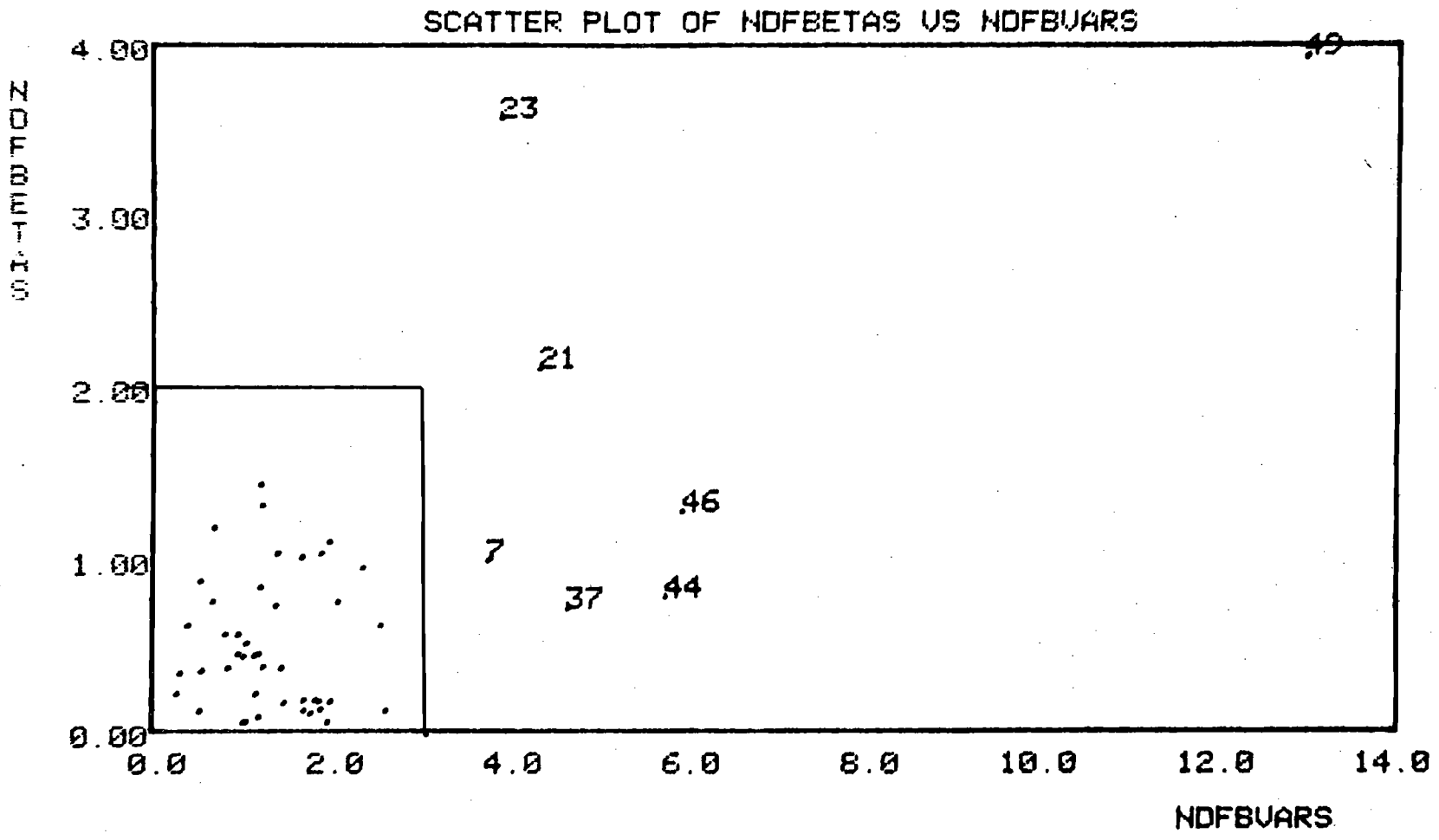

SCATTER PLOT OF NDFBETAS VS NDFBVARS
NDFBETAS
4.00
3.00
2.00
1.00
0.00
0.0
2.0
4.0
6.0
8.0
10.0
12.0
14.0
NDFBVARS
49
23
21
46
7
37
44

EXHIBIT 22

INDEX PLOT OF H

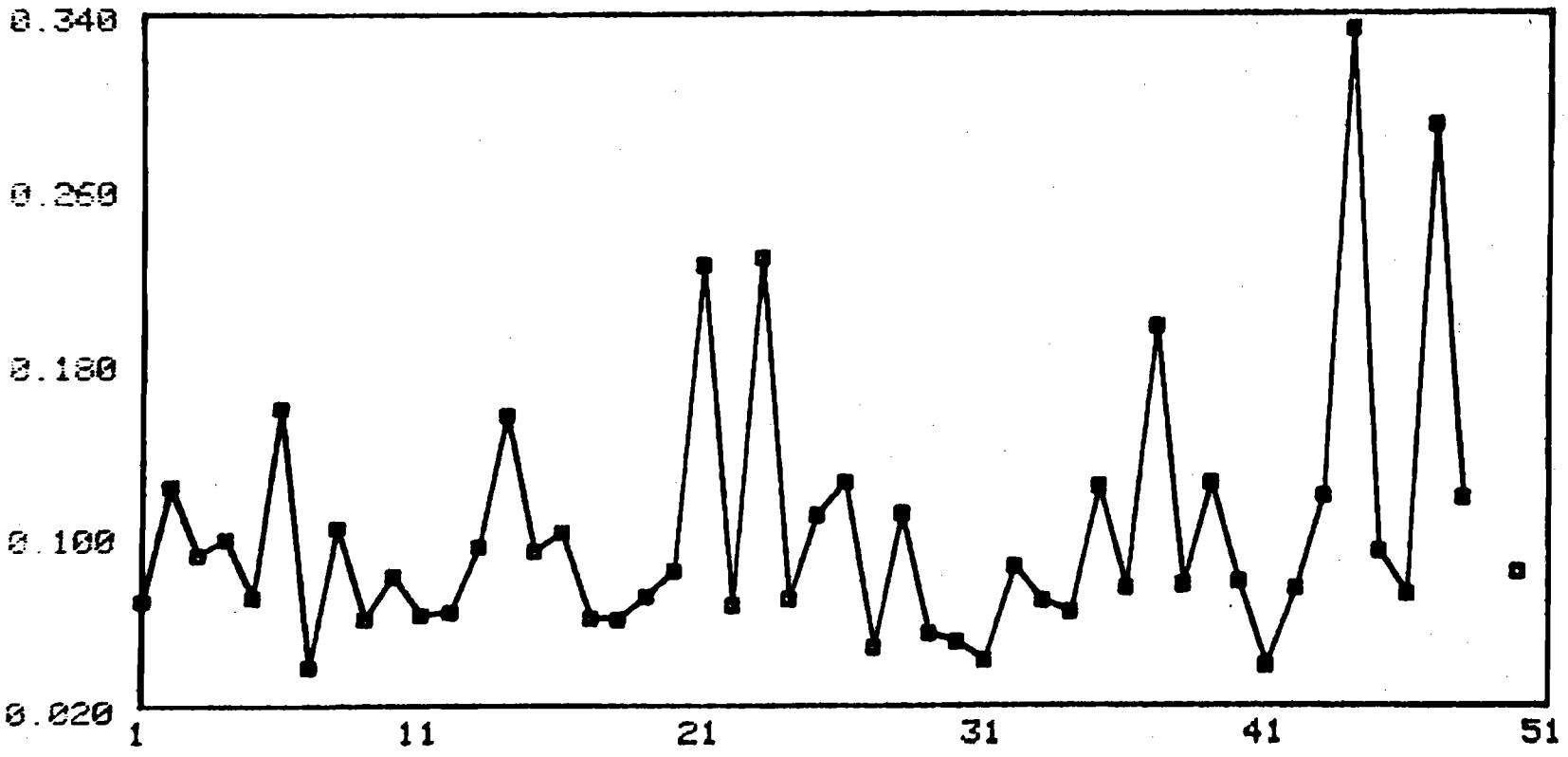


EXHIBIT 23

INDEX PLOT OF NDFBETAS