NBSIR 79-1711

Resource Requirements and Allocations in IRS' Audit Division

Karla L. Hoffman
Lambert S. Joel
Martin H. Pearl

Center for Applied Mathematics
National Engineering Laboratory
National Bureau of Standards
Washington, DC 20234

February 1979

Final

Technical Report to:

**The Audit Division
Internal Revenue Service
1111 Constitution Avenue
Washington, DC 20224**

NBSIR 79-1711

RESOURCE REQUIREMENTS AND ALLOCATIONS IN IRS' AUDIT DIVISION

Karla L. Hoffman
Lambert S. Joel
Martin H. Pearl

Center for Applied Mathematics
National Engineering Laboratory
National Bureau of Standards
Washington, DC 20234

February 1979

Final

U.S. DEPARTMENT OF COMMERCE, Juanita M. Kreps, *Secretary*

Jordan J. Baruch, *Assistant Secretary for Science and Technology*

NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Director*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Resource Requirements and Allocations in IRS' Audit Division

Karla L. Hoffman
Lambert S. Joel
Martin H. Pearl

Abstract: The Applied Mathematics Division of the National Bureau of Standards (NBS) was asked by the Internal Revenue Service under IRS order #7T2867 to "assess the validity and effectiveness of the IRS Audit Division's Long Range Plan's strategies and approaches to resource requirements and allocations". This report, which documents that assessment, (1) summarizes the examination and evaluation by NBS of IRS's current audit practices and plans for the future, and (2) presents our major conclusions and recommendations. In many cases the available information did not permit recommending specific methods with which to solve current problems. In such situations we sought to identify those areas in which we believe further research is needed and is most likely to lead to improvements over present practices.

1. Introduction

The Applied Mathematics Division of the National Bureau of Standards (NBS) was asked by the Internal Revenue Service under IRS order #7T-2867 to "assess the validity and effectiveness of the IRS Audit Division's Long Range Plan's strategies and approaches to resource requirements and allocations". This report, which documents that assessment, (1) summarizes the examination and evaluation by NBS of IRS's current audit practices and plans for the future, and (2) presents our major conclusions and recommendations. In many cases the available information did not permit recommending specific methods with which to solve current problems. In such situations we sought to identify those areas in which we believe further research is needed and is most likely to lead to improvements over present practices.

This review pertains to IRS policies and procedures. Pertinent IRS records were reviewed, data were analyzed and discussions were held with national office personnel. However, no contacts were made with "field" personnel and problems specific to them were not considered. Divergences between practice and policy, though obviously of potential significance, were beyond the scope of this study.

1.1 Role of IRS Audit Division

As stated in the IRS 1975 Long-Range Plan,[*]

---

[*] An unsigned IRS internal draft document.

"The audit mission is to encourage and achieve the highest
possible degree of voluntary compliance with the require-
ments of the internal revenue laws for the correct reporting
of income, estate, gift, employment and certain excise taxes
through implementation of appropriate programs for the
measurement of the types and degrees of compliance in
reported tax liabilities, the determination and analysis of
the reasons for noncompliance, and reduction of noncom-
pliance by maintaining optimum levels of examinations for
all classes of taxpayers."

The audit process begins when a return reaches one of ten IRS service
centers where it is checked manually for completeness and obvious
inaccuracies (e.g. unallowable deductions). The service center's
computer then checks the correctness of the taxpayer's computations.
Eventually, each return is scored by the computer using a mathematical
formula(DIF)—obtained from a Discriminant Function Analysis—which
indicates the likelihood of tax error. The returns having the highest
DIF scores are then reviewed manually, and those which are deemed to
have the highest error potential are assigned for audit.

The first two steps in the procedure just described are straightforward
and will receive little attention here. However, a major portion of
this report is concerned with how best to carry through the last step,
i.e., how to decide which tax returns are to be audited. The second
part of that decision (the manual review) is left to the judgment of
experienced personnel. It is the prior portion, the identification by
the computer of tax returns proposed for audit, which will concern us
initially.

In order to evaluate, on a continuing basis, the effectiveness of the
procedures it uses to perform its duties, the Audit Division must
incorporate into its Long-Range Plan: (1) the maintenance of an adequate
data base from which to estimate current levels of compliance; (2) the
development of optimal methods for identifying those tax returns with
high likelihood of error; and (3) means for measuring the effect on
voluntary compliance of the number and type of audits conducted. Sec-
tions 2, 3 and 4 of this report will discuss respectively the present
data base, the methodology used to identify noncomplying taxpayers,
and the influence of auditing on compliance.

In the discussions we will address how each of these major areas of
concern relates to the Audit Division's goals, as stated in the IRS
1976 Long Range Plan, of: (1) promoting a high degree of voluntary
compliance; (2) satisfying considerations of equity; and (3) maxi-
mizing the net direct yield from Audit examinations. Section 5 will
then evaluate IRS' Long Range Plan, and Section 6 will summarize our
findings.

2. The Data Base

As set down in its mission statement, a function of "Audit" is to measure degree of compliance with the law and types of deficiencies in reporting tax liability. IRS derives this information largely through the Taxpayer Compliance Measurement Program (TCMP) surveys. Randomly selected income tax returns from individuals, small corporations (assets less than $1,000,000), estate exempt organizations, and fiduciaries undergo intensive audits. Whenever an individual income tax return selected in the TCMP process has a partnership schedule attached, and that partnership has fewer than 11 partners, the partnership itself, and all partners undergo an intensive audit. (These partnership returns actually increase the number of returns in the final sample). The results of these in-depth audits (which are reviewed for completeness, fairness and accuracy) are used to determine the present levels and trends of compliance by the tax return population.

Voluntary compliance level (VCL) is measured by comparing the tax liabilities voluntarily reported (TR) by taxpayers, with the "correct" total tax liabilities (TL) as determined by Audit in the TCMP survey. Specifically,

$$\text{VCL (in \%)} = \frac{TR}{TL} \times 100$$

$$\text{and } TL = TR + TC$$

where TC = Total disclosable tax change on all returns filed by class (projected from the TCMP sample findings) as if all returns filed were audited.*

The first TCMP Survey sampled individual returns filed in 1964, and subsequent surveys have sampled returns filed in 1966, 1970, 1972, 1974, and 1977. Comparison of TCMP data for different years can reveal both shifts in general compliance levels and specific changes in tax reporting behavior. TCMP data are also used to develop and test DIF formulae. Moreover, they provide a means of estimating audit yields (average tax changes and no-change rates) at various levels of examination coverage. Without this information, IRS would have virtually no empirical basis on

---

*Until recently, the total disclosable tax change on all returns was defined to be the sum of the absolute values of all tax underpayments and tax overpayments. In the revised computations now in use, "total disclosable tax change" includes only tax underpayments. The change was apparently made for nontechnical reasons associated with bewilderment arising from identification of overpayment as "deficiencies".

which to evaluate its programs or to plan an optimal audit strategy. We applaud IRS's investment in this research.* It should be noted that the research effort itself affects compliance behavior; the Taxpayer Compliance Measurement Program is well known to sophisticated taxpayers, who are therefore aware that they have a nonzero probability of being audited no matter how successful they have been in maintaining low DIF profiles. It is also known that TCMP audits will be especially intense, and will not merely examine the most glaring line items on each return. This awareness, that a random process is part of the Audit Procedure, enhances the IRS's audit strategy in the sense that no one can ever be certain of avoidance of audit. Such uncertainty should improve the "voluntary" compliance of some taxpayers.

One of the most important uses of the TCMP data is to generate DIF formulae for the various audit classes. The next section will analyze the appropriateness of the present stratification into audit classes and evaluate the effectiveness of the current DIF formulae for these classes.

---

*This statement, which may appear patronizing, is made to emphasize our strong belief in the (frequently ignored) importance of basing decisions on appropriate data.

3. Audit Classes and Their DIF Formulas

The present audit classes, shown in Table 1 following, are described in the July 1976 draft document "Audit Planning and Resource Allocation" prepared by the IRS Audit Division:

> "For planning and program management purposes, both individual and corporation returns are stratified into classes according to certain common characteristics, i.e., amount of adjusted gross income (AGI) (plus the characteristics of standard or itemized deductions in low AGI classes and the presence or absence of business schedules) for individual returns, and asset size for corporation returns. Presently, eight classes are used in the individual return area and eleven classes in the corporation return area. The detailed class breakdown is shown on Chart 13, Appendix 10--Audit Strategy For the Long-Range Plan.
>
> Tax returns as a whole constitute a greatly diverse population, and are stratified to form relatively less heterogeneous groups to facilitate planning and program management. For example, simple types of returns require relatively low level of audit skills and lower average time of examination as compared with the complex ones. Also, it has been found that there is a reasonably good correlation between the average size of tax error and the income level (or asset level in the case of corporation returns), type of deductions and complexity of tax issues involved. Thus an appropriate stratification helps to enhance the effectiveness of planning and program management by taking advantage of the cost and benefit differentials among different classes of returns.
>
> Five years ago, a study was conducted to test a return selection procedure for individual returns without any class stratification at all. Using TCMP data, a so-called one-class DIF formula was developed for all sample individual returns. As expected, the formula would select primarily high-income returns for audit. For a given level of audit resources, it would mean a lower level of overall audit coverage since examinations of high-income, complex returns take more examination time per return. Moreover, with such a workload Audit would need only technical personnel with high-level accounting skills, who are in relatively short supply. Furthermore, the highly one-sided concentration on high-income taxpayers and the corollary absence of coverage elsewhere certainly causes concern about adverse impact on voluntary compliance of the great numbers of taxpayers. These observations also indicate that a very gross scheme of class stratification would tend to reduce planning and program management effectiveness.

For pragmatic reasons, however, we cannot afford to have too many audit classes, even though theoretically the greater the number the more homogeneous are the elements within each class for fuller revelation of the co-benefit differentials between classes. This is because too detailed a stratification scheme would require a similarly detailed programming of operations of all field units, not to mention the corollary reporting and data processing requirements. Additionally, increasing the number of classes significantly will cause a further fragmentation of the data base or a very "thin" or limited data base for each planning unit, which would tend to result in highly unreliable planning estimates or projections.

The preceding discussion cannot be construed as an argument that the current audit class stratification scheme is the optimal one. It should be reexamined from time to time. Limited studies of the reclassification problem have been conducted in the past, e.g., classifying business returns on the basis of gross receipts instead of AGI, but these studies have not produced definitive findings except that a splitting of the medium non-business class into two income subclasses would be helpful. However, in view of the recent inflation phenomena, reclassification efforts should be resurrected to test the idea of "indexing" the income class breaks (or putting the class breaks on a constant dollar basis). Similarly, perhaps it is worthwhile to split the medium-income non-business class (MNB) into standard and itemized deductions subclasses to overcome the apparent anomaly of having a planned coverage for low-income nonbusiness (itemized deduction) class higher than the current MNB class (with both standard and itemized deduction returns)."

We concur with the Audit Division's general philosophy of stratifying returns into audit classes to form less heterogeneous groups, and agree that a very large number of classes would cause many administrative problems and require a large expenditure being directed toward TCMP surveys. We also agree that the existing stratification should be reexamined regularly. Section 3.3 suggests a number of topics for inclusion in such a reappraisal.

3.1 Evaluation of the DIF Process

To test the quality of IRS's stratification of returns into audit classes, we must examine how well the corresponding DIF formulations for each class succeed in differentiating those returns with a high probability of significant tax change from those with low probability. To start, we can examine how well the DIF scores for a set of tax returns (chosen randomly by a TCMP survey and presumably covering a broad range in degree of compliance), correspond to the disclosable tax changes for the same set of returns. Disclosable tax change is

6

defined to be the difference between tax liability voluntarily reported on tax forms and the "correct" total tax liability as determined by Audit. "Tax change" in this report will always refer to the absolute value of the changes, i.e. to overstatement as well as understatement of tax liability. Although, in principle, discriminant analysis is meant merely to separate a population into a fixed number of classes (in this case two: compliers and noncompliers) the mechanism used for this separation should yield "good" correlations with tax-change sizes when the DIF is effective. It has been standard practice by IRS to use the DIF score for ordering the returns.

The Spearman rank correlation coefficient and the Pearson correlation coefficient are two measures of the degree of relationship of two sets of data. Since the Spearman test deals only with ranking, it is less likely to distort the sense of the DIF (which attempts to discriminate compliers from noncompliers rather than predict the degree of non-compliance) than is a test based on explicit numerical values, such as the Pearson test.

At NBS, the DIF scores and the tax changes within each audit class were both ranked from highest to lowest. The Spearman correlation coefficient produces an index of the agreement between these two sets of rankings. Table 1 lists the correlation coefficients for some audit classes. They were lower than had been expected. A ranking procedure correlating poorly with a target ranking over an entire set might still be quite useful if the correlation at the high end of the target ranking were good. We therefore proceeded to check whether the discriminant function might perform better for those returns within each class which have high DIF scores. The Spearman rank correlation coefficients for the actual coverages planned for 1977 were calculated, and are also displayed in Table 1 below. Contrary to our hopes, the correlation coefficients for the actual coverages planned for 1977 were lower than those for the entire audit class.

7

TABLE 1

SPEARMAN RANK CORRELATION COEFFICIENTS (ADJUSTED FOR TIES)

MEASURING RELATIONSHIP OF TAX CHANGE TO DIF SCORE

TCMP SAMPLE OF RETURNS FILED IN 1970

	Entire TCMP Sample (Unweighted)	For Coverage Levels Planned for 1977 Based on TCMP Data
INDIVIDUAL:		
Under $10,000 Standard	.54655	.43327 (.6%)[*]
Itemized	.36451	.28973 (4.2%)
$10,000-Under $15,000	.37254	.23364 (2.0%)
$15,000-Under $50,000	.27953	.14385 (2.2%)
$50,000 and Over	.23924	.18944 (11.0%)
NONFARM BUSINESS:		
Under $10,000	.19102	.12733 (3.6%)
$10,000-Under $30,000	.34770	.32190 (2.5%)
$30,000 and Over	.31871	.13131 (10.2%)

[*]
 The numbers in parentheses indicate the coverage levels for each audit
class. For example, the coverage of individual returns with AGI or
$50,000 or more would consist of the 11.0% with highest DIF scores. This
table excludes taxpayers with income from farms, because those data
were not supplied to NBS.

These results were discussed with IRS personnel, who then indicated that
they too had calculated correlation coefficients for TCMP data based on
returns filed in 1966. They first aggregated the data by dividing each
audit class into between 15 and 31 intervals, depending on the class.
For each subdivision the mean (average) tax change and median DIF score
were calculated. The Pearson (linear) correlation coefficient obtained
by comparing these interval measurements yielded significantly higher
results, as is shown in Table 2.

TABLE 2

PEARSON CORRELATION COEFFICIENTS BASED ON AGGREGATED DATA

MEASURING RELATIONSHIP OF AVERAGE TAX CHANGE TO AVERAGE DIF SCORE

TCMP SAMPLE OF RETURNS FILED IN 1966

		Aggregated into the Following No. of Intervals	Pearson Coefficient
INDIVIDUAL:			
Under $10,000	Standard	31	.8553
Under $10,000	Itemized	23	.9449
$10,000-$50,000		29	.8504
Over $50,000		25	.8207
NONFARM BUSINESS:			
Under $10,000		21	.8194
$10,000-$30,000		28	.8908
Over $30,000		27	.8497
FARM BUSINESS:			
Under $10,000		23	.7601
$10,000-$30,000		15	.8198
Over $30,000		26	.5991

The results in Tables 1 and 2 cannot meaningfully be compared. The two tables use different data, different correlation coefficients, and one uses aggregated data whereas the other does not. For this reason we reexamined the data used in compiling Table 1, and for each audit class aggregated the data so that approximately 4% of the population was contained in each of 25 intervals. A second aggregation was also computed which divided the class into 100 intervals each containing approximately 1% of the population. For each subdivision the average tax change and median DIF score was calculated and a Spearman Rank correlation coefficient obtained. The results, presented in Table 3, are discussed in the next paragraphs.

The correlations using aggregated data are seen to be higher than when unaggregated data are used. This is to be expected with either the Pearson or Spearman correlation coefficients. The aggregation process smoothes the data, thereby diminishing the effect of outlying points and increasing the correlation.

The Long Range Planning process determines the percentage level (to one decimal place) of audit coverage for each class, based on expected direct yield from audit. The relationship between expected yield and audit coverage is defined in terms of a discriminant function whose

9

values are used to separate the levels of coverage. If the process of ranking returns is to merit the adjective "efficient", then although some deviations from the ranking by tax change would be tolerable for the unaggregated population, certainly the ranking for the population aggregated into 4% chunks (thus reducing many thousands of members to just 25) should be perfect within the feasible audit coverage levels (i.e. the average DIF score for the highest 4% DIF scores should be greater than the average DIF score for the next 4%). Plots of the data exhibited in Appendix C highlight the ranking deficiencies which can be deduced from Table 3 below.

TABLE 3

SPEARMAN RANK CORRELATION COEFFICIENTS (ADJUSTED FOR TIES)

MEASURING RELATIONSHIP OF AVERAGE TAX CHANGE TO AVERAGE DIF SCORE

TCMP SAMPLE OF RETURNS FILED IN 1970

	Aggregated into 25 intervals (4% in each interval)	Aggregated into 100 intervals (1% in each interval)
INDIVIDUAL		
Under $10,000 Standard	.95712	.83684
Itemized	.84911	.84253
$10,000-Under $15,000	.83385	.69865
$15,000-Under $50,000	.93769	.75154
$50,000 and over	.95077	.80961
NONFARM BUSINESS		
Under $10,000	.84231	.43885
$10,000-Under $30,000	.96385	.77699
$30,000 and Over	.95846	.80000

It is quite possible for two returns to be identical "on face" (so that any automated classification scheme would fail to separate them), and yet disclose widely differing tax changes after audit. Often a return which looks suspect if only its line items are examined, has supporting documentation attached which establishes the admissibility of the questionable items. Such distinctions are made after the DIF process, by classifiers who screen manually all returns having high DIF scores and choose among those returns the ones they believe have the greatest potential for tax change. It will be convenient for us to refer to the phenomenon of returns which cannot be separarated by any automated classification scheme as "noise".

We must conclude from the correlation coefficients presented in Tables 1, 2 and 3 that the present DIF process does not produce a good ordering of those returns with high tax change. In the presence of substantial

10

noise, even the "best" DIF formula would not be able to produce "good" correlations. Thus it is quite possible that high correlations on unaggregated data may not be obtainable. At this stage of our study we are not able to determine how much the low correlations can be attributed to large random error, how much to inappropriate stratification into audit classes and how much to poor DIF formulae.

Insight as to how far from perfect the present methodology is can be obtained by comparing DIF results to those from random and from perfect selection. Table 4 shows the results of two comparisons by IRS of the average tax change expected under the DIF system (at a specified audit coverage level) with the average tax change under random selection (represented by the total TCMP sample), and also with that obtainable from auditing those returns with the highest tax change (representing a perfect selection scheme).

TABLE 4

COMPARISON OF DIF, RANDOM AND PERFECT DISCRIMINATION

Audit class	Percent of audit coverage (note a)	Average tax change per return		
		DIF	Random selection	Perfect selection
Nonbusiness:				
Low	1.3	$ 230	$ 36	$ 745
Medium	1.5	802	93	1,800
High	11.7	9,460	2,178	15,312
Business:				
Low	2.0	940	230	3,265
Medium	2.4	2,755	345	5,512
High	12.6	5,781	1,662	10,590

a/ This is the approximate percentage of returns by class that IRS audits in a fiscal year.

Clearly, although the DIF system is superior to a totally random selection process, the comparison between the DIF results and perfect selection indicates considerable room for improvement. That is, even though it is intrinsically impossible to create perfect DIF selection formulae, improvements in the present system may be possible. The weakness of the present system may not be due entirely to random error. Making available to the discriminant-function analysis process a greater body of information with which to distinguish compliers from noncompliers, creating more homogeneous audit classes and/or improving the discriminant-function methodology, might permit a scoring system that ranks tax returns more accurately.

11

Cursory analysis identifies three potential sources of supplementary information for discriminant analysis:

1.   There exists the possibility of requesting more information about a taxpayer than is currently available from the tax return (e.g., age, education, home ownership, etc.) and incorporating some or all of this information into the DIF formulae. However caution is necessary when moving in this direction, since "privacy" questions may arise.

2.   Veteran tax-return classifiers could be consulted to determine on what basis they conclude that a tax return should be audited. Although these judgments may be largely intuitive, it may very well be possible to formalize some of the criteria used by these classifiers and to incorporate some of this information into the DIF formulae.*

3.   Tables of supplementary information should be available as input into the Discriminant Function Analysis as a basis for making comparisons with the tax returns. For example, the expected rate of return or the ratio of inventory to gross receipts for particular broad categories of business (the first two digits of the SIC code), as given by Department of Commerce figures, can be compared with the corresponding figure for an individual business of that type as reported on its tax form. Thus a 10% rate of return might be completely within reason for one type of business and yet completely anomalous for another type.

3.2 Exploration of the Discriminant Function Analysis Methodology

The methodology of the discriminant function analysis underlying the DIF scores was examined in order to determine whether deficiencies in the method could account for the observed poor performances and whether improvements in the process could be effected by a change in the audit class definitions.

The methodology employed by IRS entails the use of two discriminant functions, one developed in-house and the other developed independently by an outside contractor. Both functions are applied to TCMP sample returns in an audit class, and the one yielding the better results (higher average tax change and lower no change rate, for the same audit coverage level) is the one chosen which will be applied to all returns for that class in subsequent years. The methodology of the outside contractor was not fully documented for IRS because of proprietary considerations.

---
*
 We have been informed that IRS is currently making a study in which TCMP returns are subjected (before the audit) to the customary manual screening process by classifiers. This will provide some measure of the discriminating power of the classifier and could serve to begin the formalization of some of the classification procedures.

The methodology of the in-house development was described to us in detail and, in theory, seems appropriate. It consists of choosing approximately 200 income tax variables (data available from income tax returns) to be used to differentiate those returns with significant tax change potential from those with little or no potential. The tax returns in the data base (TCMP tax returns) were grouped into three intervals:

Subset 1: No tax change or less than $X,
Subset 2: Tax change between $X and $Y,
Subset 3: Tax change greater than $Y.

The range of each variable is divided into intervals (usually between six and ten). Then the frequency distribution of the variable over the returns in subset 1 is compared with its frequency distribution over subset 3. Those variables for which there is little difference in distribution between subsets 1 and 3 are eliminated, as are variables for which there are insufficient data. Subsets 1 and 3 are chosen in the analysis in order to differentiate the "no change" population from that with significant tax changes (the value Y varied among classes, but X was always set at $25).

After this preliminary screening, approximately 100 variables remain. Likelihood ratios for each interval (the ratio of relative frequency of subset 1 to relative frequency of subset 3) for each variable are calculated and used as preliminary weights in the discriminant function. The original likehood functions (preliminary weights) are then modified using the covariance matrix to obtain the final weight or contribution of that variable.

A linear discriminant function was developed. The methodology is commonly justified in terms of assumptions of normality and identical covariance matrices. Since the latter assumption was not met, the data of Subset 1 and 3 were pooled, (a standard approach), and a single covariance matrix calculated. It was hoped that this matrix obtained by pooling the data was a good approximation to the underlying population covariance matrix.

All 100 of the screen-surviving variables are considered in the discriminant function analysis. A weight for each variable is assigned in the usual way by this process; the three variables found to contribute the least to separating the two populations are then removed and another discriminant function analysis performed. The three variables found to contribute the least are again dropped and the process continues until a discriminant function formula with 25 variables has been calculated. Everything in the analysis up to this point is totally automated, and only the final 25 variables (and their correlation matrix) appear in the computer print-out.

IRS staff examine this correlation matrix, deciding which (if either) of pairs of variables having high correlation (greater than 0.7) will be

removed. At this point, they try a number of alternative modifications by forcing variables either into or out of the discriminant function. The formula having the "best" results in terms of no-change rate and average tax change for the same coverage level, is the one chosen.

The contractor-developed DIF development involved an iterative procedure in which the contribution of each return characteristic is measured in terms of the incremental information it provides in discriminating between the two subpopulations. This system, unlike the in-house system, uses all the tax change information including the "gray area" of Subset 2. It considers the little-or-no-change area to include those returns with tax change less than or equal to Y, and the profitable-to-audit population to be those returns with tax changes greater than Y. The methodology of this discriminant approach allows unequal variances in the two subpopulations. Therefore two covariance matrices (rather than the single pooled matrix used in the in-house analysis) are provided.

The results of applying each of these two systems to the data base are compared, and the one yielding "better results (higher average tax change and lower no change rate, for the same audit coverage level) is the one chosen. Having two methodologies available allows IRS the flexibility of choosing an approach which best reflects tax-payer behavior in a particular audit class. This ability to switch approaches also makes it more difficult for someone outside IRS to be able to "break the code" and accurately estimate what a taxpayer's DIF scores might be.

In principle, this automated process assures that each of the over 70 million individual tax returns and each of the over 1.3 million corporate returns with assets under one million dollars is screened. Of those returns, only those believed to have the greatest change potential are then selected for audit. The objective of equity is thus advanced because the DIF methodology again, in principle, assures that scarce enforcement resources are directed toward those taxpayers with the greatest expectation of noncompliance, while assuring that those taxpayers expected to be in substantial compliance will not be asked to incur the additional burdens of being audited. The process should also promote the confidence of taxpayers in the equitable and evenhanded administration of the tax laws.

It is easily seen from Table 4 that the current DIF system yields results well below those which would be obtained if the selection process were perfect. While perfect selection cannot be expected and the current DIF methodology seems quite reasonable, we nevertheless offer a few alternative procedures which we believe are worth exploring for potentially producing a better separation of tax returns according to high probability of error.

In a very lucid paper [4] (whose details we do not repeat here), Kendall describes a discrimination procedure which makes no assumptions

14

about the probability distribution of the populations. This "distribution-free" methodology, unlike standard discriminant analysis, does not separate the discriminated populations by a simple curve (e.g. line, plane) but by an irregularly shaped boundary. It requires less computation than standard discriminant analysis but it is logically more complex and does not furnish a single numerical "score".

As an alternative to any discriminant techniques which separate rather than order populations, we suggest that methods based on the notion of regression be evaluated. These methods have the conceptual advantage that a function is chosen which is a direct measure of the expected tax change on a return and is therefore designed to rank returns.

There are two alternative methodologies to explore within this framework. The first approach requires a preprocessing of the data using the mathematical procedure known as principal component analysis to remove collinearities in the data, followed by a standard regression analysis. The second approach is based on heuristic pattern searching procedures. The rationale for this approach is that multi-valued functional relationships will escape standard regression techniques whereas this one has the possibility of disclosing them. (For instance, a response pattern in the form of a circle will show up as "noise" in classical linear regression.) These procedures allow the data analyst to graphically draw in two dimensions meaningful pictures of higher dimensional data, and can be used for still another kind of separation (as in discriminant analysis) as well as a substitute method for regression. An exposition of these procedures can be found in a paper by Dixon and Jennrich.[2]

Since the methodology used to determine which returns have a high probability of audit is the backbone of any audit strategy, we believe a thorough study of these and other approaches is relevant to IRS' Long Range Plan.

3.3 Audit Class Definitions

As already discussed, IRS stratifies returns into classes for purposes of planning and program management. All decisions are then made on a class by class basis. For each class, a discriminant function formula is determined, audit yields and costs of audit are projected and the voluntary compliance of each class is estimated. Comparisons are then made among classes to determine the proportion of audit resources each class will receive. The method used for this determination will be discussed in Section 5. Different definitions of audit classes (i.e., class boundaries) will normally yield different results. Since the limited studies performed previously have not confirmed the preferability of the present audit class structure, we urge continued investigation in this area.

IRS is presently initiating a new study to evaluate the consequences of using alternative definitions of audit classes. We believe that this study should be carried out on a scale allowing a broad spectrum of alternative definitions to be tested. Going beyond an examination of the consequences of altering the boundaries of the existing classes, or of putting the classes on a constant dollar basis, it should consider a variety of possibilities with the objective of obtaining definitive answers on the benefits and costs of each. The possibilities to be addressed include defining classes on the basis of:

- geographical location

- specific line items or groups of line items on the return

- industry groups (e.g., production, service, financial, whole-
    sale, retail)

- gross receipts (for business classes)

- assignment to the nonbusiness class of individuals with income
    primarily from nonbusiness sources

- occupational groups (professional, businessman, white collar,
    blue collar, etc.)

- personal traits (single or married, number of dependents, etc.)

- self employed vs. wage earners.

Consideration should be given to developing a separate DIF score for each of the most important schedules on the tax return. Actually, a second classification will probably also be necessary. That is, a number of DIF formulae for schedule C might need to be created (e.g., Schedule C with gross receipts under $10,000; Schedule C with gross receipts $10,000-$100,000; and Schedule C with gross receipts over $100,000). By generating DIF scores for such specific classifications, one is assured that the audit classes will be more homogeneous.

Once these scores are generated, two approaches can be taken. The first of these would generate a function which integrates into a single score the results from the DIF scores on each schedule of the return, and thereby determines the returns most promising for audit.

The second approach would treat the schedules of the return as separate entities in the planning process. Audit resources would then be direc- ted towards identifying those partnership returns, business schedules, 1040 tax forms, etc. with the greatest probability for error. Under this approach the number of entities in the audit process would be substantially increased, so that the approach would be useful only if the corresponding yield from audit were significantly greater. This ap-

proach would also alter manpower allocations, since it considers portions of a return as separate entities.

The idea of generating DIF formulae for each major schedule arose from an examination of the compliance behavior of present audit classes. The low business class (individual returns with under $10,000 AGI and filing schedule C) exhibits an extremely low voluntary compliance level (63.5%). We suspected that examining the distribution of errors, by line item, might highlight approaches to improving voluntary compliance. Analysis showed that a large proportion of returns in this class require filling-out a number of complicated schedules. Twenty percent of the returns in this class showed income from rents, with a corresponding expected change in taxable income of $133 million (based on TCMP Sample data projected for the entire class as if all returns were audited); 12% showed income from miscellaneous sources (with expected changes in taxable income totalling $68 million), 4% from farms (with $41 million in changes in taxable income) and 4% from partnerships (with a corresponding taxable income change of $109 million). These data indicate that a formula for "potential for audit" for this class should refer at least to Schedule E, the partnership form, and the farm schedules, in addition to the Schedule C and the 1040 form itself, to be able to identify those returns with a high probability of error. It seems unlikely that a DIF formula with relatively few variables (at most 25) could capture sufficient information on a population as nonhomogeneous as the low business class. Moreover, one would expect that individuals with business or farm income whose adjusted gross income is greater than $10,000 would be filing returns at least as complicated as those discussed above. Thus, generating DIF formulae for each major business schedule may highlight erroneous tax returns overlooked otherwise.

These suggestions should not be construed as recommendations for an immediate redefinition of audit classes, but rather as identifying promising ideas to be analyzed and evaluated in a research project adequate to determine the costs and benefits of various new approaches to stratifying taxpayers into classes. The study will have to examine a significant amount of data over a period of years to be sure that any change in strategy is justified, and which is best. Previous studies (cf. page 5) have been too small-scale to provide sufficient evidence for choosing one approach over another.

Generally speaking, stratification of tax returns into audit classes has been accomplished to date by a methodology which is almost totally intuitive. The suggestions above have been formulated in the context of extending and refining these intuitive notions. But a basically different approach is also possible:

Although the intuitive stratification concepts appear fundamentally sound, palpable difficulties have surfaced in attempts to specify and measure the most appropriate boundaries for the classes. The source of these difficulties is the lack of a definition of within-class homogeneity which can be related to the prediction of tax change. These

difficulties suggest that some investigation of "formal" methods for stratification might be profitable.

A large number of techniques for stratification, with supporting theory, have appeared in the mathematical-statistical literature under the heading "cluster analysis". This collection of procedures has been developed largely in response to problems of classification in the life sciences and behavioral sciences, for example in medical diagnosis (classifying heart patients, for instance, on the basis of analysis of a body of past histories and current cardiograms). At least one area of extensive application, namely market research (selection of target markets for particular products and projection of sales within markets) appears to be closely analogous to stratification into audit classes followed by projection of tax changes; in fact, some of the same variables could conceivably be employed in both of these two-stage processes.

We note that the clustering or grouping problem of defining audit classes is complementary to the discrimination problem within a class. The mathematical descriptions of the two problems are similar and, in fact, "discriminant analysis" is frequently subsumed under "cluster analysis" in modern treatments of the theory of multivariate analysis. (For a more detailed description of cluster analysis, see [1] and [3]. The book by Anderberg is lucid and contains many examples; its introduction in particular is an excellent exposition. The monograph by Duran and Odell is short but fairly technical, having been written for a mathematical reader. We list it mainly because of its bibliography of 407 titles.)

The last two sections have discussed two major aspects of the Audit Division's mission, namely, measuring existing compliance and determining procedures to identify noncomplying taxpayers. In addition, the Audit Division must also conduct a sufficient number of audits to assure high compliance. The next section will briefly discuss the topic of compliance and the factors which affect it.

4. Compliance

It is the mission of the Service as a whole, and of the Audit Division specifically, to manage their programs in a manner which fosters a high level of voluntary compliance by the taxpayers. Until recently, the Audit Division had postulated the use of the "Gompertz curve" (an S-shaped curve widely used in actuarial analysis) to represent the relationship between audit coverage and voluntary compliance.

This function embodies the assumption of a close relationship between prior-year audit coverage (the percent of the population audited) and the compliance levels of the following year. It depicts decreasing marginal growth in compliance: as audit coverage increases, voluntary compliance increases at a diminishing rate, and will, at some level of coverage, level off at a maximum attainable level. The maximum attainable level will probably be below 100% since certain taxpayers will be reluctant, or possibly unable, to change their behavior even if a high probability of audit exists.

IRS has discontinued using this model for a number of reasons:

1. The model entails the assumption that audit coverage is the factor having the greatest effect on compliance. Although a limited number of studies have illustrated that audit does have a positive effect on compliance, many other factors also alter taxpayer behavior, and the Gompertz curve does not reflect these factors.

2. The model needs, as input, a numerical value for the maximum attainable level of compliance. For large corporations (whose audit coverage is 100%) voluntary compliance levels are not greater than 94%, so that there is empirical evidence that compliance will converge to less than 100%. But for most audit classes, no data exist from which to predict this asymptotic level.*

3. The number of data points at IRS' disposal for estimating the three parameters of this nonlinear curve is at most four, which is insufficient given the uncertainties in the data. This is because TCMP data which yield the necessary empirical information on voluntary compliance are taken only every few years, and the program has not been in existence long enough to produce more data.

Having discarded use of the Gompertz curve which had been used to predict audit's effect on voluntary compliance, IRS has no formulated alternative model to employ in its planning process. No theory exists to provide the mathematical form of a more suitable functional relationship.

_____

*For some audit classes, a small fraction of the entire TCMP sample of 1971 individual taxpayers had also been audited in the prior TCMP cycle. Although compliance did improve for those taxpayers "hit twice", this sample is too small to use in determining the asymptotic level.

IRS has been aware of the need for better understanding of the factors which affect compliance, and a number of studies have addressed this question. In a 1966 study, researchers from the National Opinion Research Center [5] queried individuals who filed personal income tax returns on their perceptions of the audit process. They asked people to estimate the possibility that they would be audited, and what changes in the tax returns they would expect to increase or decrease their chances of being audited. They also solicited taxpayers' opinions on the fairness of the audit process and what types of punishments they believed to be appropriate to taxpayers who violated the law, either knowingly or unknowingly. This type of survey gives IRS information about how taxpayers perceive the audit process. In modelling taxpayer behavior, an understanding of these perceptions is necessary.

Another study performed in 1966 and described in the Chicago Law Review [6] examined the effects of legal sanctions on taxpayer behavior. It questioned the degree to which legal sanctions and appeals to taxpayer conscience affect taxpayer behavior. The authors concluded that although legal sanctions increase willingness to comply for certain classes of individuals, they have negative effects on others.   Alternatively, appeals to conscience may have greater effects on certain groups of individuals than do threats. These preliminary results suggest that IRS should study further what positive approaches can be taken to affect taxpayer behavior.

In a third study, performed in-house and entitled "A Cross Section Regression Model of Audit and Non-Audit Factors Effecting Taxpayer Complian e (RARC 7.2)", the authors disaggregated the IRS master file into five -digit ZIP Code areas and attempted to regress IRS policy variables (audit coverage, collection activity, return complexity) and external variables (unemployment rate, average AGI, educational level, population density, etc.) against voluntary compliance. Unfortunately, the study was terminated when it was found that the data available at that time did not permit accurate measurement of voluntary compliance at the three and five-digit ZIP Code level (three-digit codes used in rural areas, five-digit codes in urban areas). We commend the concept underlying this attempt and believe that, if in the future it is possible either to aggregate more than one TCMP Sample in order to obtain estimates of compliance at regional levels or to design future TCMP Samples in a manner allowing estimates of compliance by region, then a study similar to the one described above should again be undertaken. As the report states,

> "the outcomes of such a study could develop information of the following types:
>
> .01 a.   For the first time, a valid voluntary compliance audit coverage relationship may be measured.
>
> b.   In a similar vein it may be possible to determine the impact of Intelligence activity on voluntary compliance.

c.   The effects of return complexity and education on
            voluntary compliance may be isolated.
       This type of information not only increases our general
       understanding of the tax administration problem but it is also
       necessary to help secure the resources needed to design and
       operate an optimal tax administration system.
  .02 The project also has potential for real, short-term pay-offs
       by isolating geographic pockets of noncompliance. . .
  .03 Knowledge of this sort helps pinpoint not only tax administra-
       tion problems, but also basic social, economic, and political
       problems. This project, with its broad criminal-sociological-
       economic approach, should throw some light on whether the
       traditional enforcement approach or a more 'enlightened'
       taxpayer service and education approach would be more effec-
       tive in solving such basic problems as how to raise voluntary
       compliance in ghetto areas."

The studies discussed above are designed to determine the factors, both
internal to IRS and external socio-economic, which effect compliance
behavior. We believe that more studies of this nature should be under-
taken, and a list of other possible approaches follows:

  1.   Surveying taxpayers to monitor their perceptions of
       IRS activities, their ability to compute their taxes accur-
       ately, and their actions taken in attempts to comply with the
       law (e.g., withholding more than is prescribed).

  2.   Controlled studies to evaluate the effects on voluntary
       compliance of media campaigns, of the distribution of
       educational materials, of IRS' telephone-answering tax
       advice service, and of other similar IRS activities.

  3.   A study which attempts to model taxpayer behavior by
       using a game-theoretic approach. A detailed description
       of some of the data necessary to perform such modeling,
       and of the answers which this approach might supply, is
       given in Appendix A.

The above recommendations do not necessarily pertain to the Audit
Division's responsibility directly, but rather to the entire IRS
mission. We include these recommendations here because actions based on
the output of such research efforts will have direct and essential
effects upon the Audit Strategy.

Research efforts which relate more directly to the Audit function will
now be discussed. In a study recently completed by IRS, the effect of
auditing a taxpayer's return on his future compliance was examined.
Researchers analyzed 2,169 returns examined in both Cycle 3 (1969
returns filed in 1970) and Cycle 4 (1971 returns filed in 1972) TCMP
samples.

Three classes were examined: low nonbusiness (less than $10,000 AGI
with itemized deductions), medium nonbusiness ($10,000-$50,000 AGI with
itemized deductions) and low business (less than $10,000 AGI). In all
three classes examined, compliance increased after audit. A similar
study, which examined effects on taxpayers who had previously been
subjected to a non TCMP sample audit, also found increased compliance in
the returns of previously audited taxpayers.

No studies have been made to test the "ripple" effect of audit, that is,
the effect that an audit will have on persons who have some contact with
the audited taxpayer. Even though studies to determine "ripple" effects
are difficult to design, and unlikely to produce definitive results,
such research efforts can produce useful insights and should be pursued.
When a district or region initiates audits of a certain class of indi-
viduals, (e.g., those in a specific profession or business), the
"ripple" effect on others in that profession or industry can be eval-
uated by subsequent auditing or by examining changes in the reporting
characteristics of tax returns for that class. Another topic calling
for investigation is the rate of decay of the increased compliance in
the second and subsequent years following an audit.

The IRS Audit Division has often asserted, among other justifications
for its manpower requirements, that audits do have a positive "ripple"
effect and that therefore an increase of any significant size in
coverage will have a manifold impact on voluntary compliance. The
assumption certainly appears reasonable, but studies to test this
hypothesis should be undertaken.

Examination of the previous studies initiated by IRS has shown that very
little is as yet known about the factors which affect compliance and how
these factors interact. Our recommendations are aimed at improving this
situation. Since the mission of the IRS is "to encourage and achieve
the highest possible degree of voluntary compliance with the tax laws
and regulations," substantial research addressing how both audit and
nonaudit factors influence taxpayer behavior is justified.

We have, in the past few sections, discussed the major components of the
audit process: the Taxpayer Compliance Measurement Program, the Dis-
criminant Function Analysis methodology, the stratification of taxpayers
into Audit Classes, and Compliance. The next section will briefly
summarize the 1976 and 1977 Audit Long-Range Plans, explain how each of
these components fits into those plans, and give an overall evaluation.

5. The Long Range Plan (LRP)

The first part of this section will address the income and estate tax program areas, where the bulk of the analysis and planning effort is concentrated and for which the largest body of pertinent data exists. Some general comments will then be made about the other tax areas: fiduciary, gift, excise and employment tax returns.

Since 1970, IRS has been using a so-called "balanced strategy" whose objective is "maximizing total net direct yield subject to the constraint of maintaining certain minimum levels of voluntary compliance in the known low compliance classes." Under the strategy, sufficient resources are placed in low complying classes to assure that compliance within five years will be raised to a specified level (85%, in the 1976 LRP). The remaining resources are then allocated among all classes to maximize net yield, constrained again to assure that if sufficient resources are available, they are placed in each class so that voluntary compliance in the complying classes does not fall below a specified level (90%, in the 1976 LRP).

It is a well-known theorem of mathematical economics that allocating resources so as to equalize the marginal-yield to marginal-cost ratios of all classes under consideration will achieve the same allocation as that of maximizing direct net yield. Although, given current cost and yield data, pursuing this policy will result in the allocation of some resources to every audit class, some low complying classes will receive very small allocations of resources. The strategy discussed above is therefore modified to one of <u>constrained</u> maximization which forces additional resources into the low complying classes. Under both strategies, taxpayers in the higher income (asset) classes receive high coverage because they have the greatest possibility for significant tax errors. The constrained approach promotes equity additionally in the sense that taxpayers in low complying classes are not overlooked simply because the cost of audit is high relative to the likely yield. Table 5 (pg. 24) presents the 1982 proposed coverage levels based on the balanced strategy used in the 1976 Long Range Plan.

This balanced-strategy approach requires a methodology for estimating the examination coverage needed to reach or maintain a specified level of voluntary compliance. The so-called Gompertz curve is the mathematical expression which had been previously used to model the compliance-coverage relationship. Section 4 discusses the reasons for IRS' abandonment of this procedure. Unfortunately, at present no better model exists with which to represent this relationship.

Because of this lack of information, the new plan being initiated this year takes the approach of improving compliance in low complying classes and maximizing net yield (by equalizing the ratio of marginal yield to marginal cost in each audit class) in all classes. This strategy requires that sufficient funds be allocated to low complying classes so that the gap between that which was voluntarily reported and that which

TABLE 5

PROGRAM AND CLASS COVERAGES IN FY 1982 AND THE FY 1977
TENTATIVE EXAMINATION PLAN

	FY 1977 (Tentative Plan)[1]	FY 1982 (1976 LRP)[2]
INCOME		
INDIVIDUALS (Audit Class)		
NONBUSINESS		
UNDER $10,000		
Standard	0.6% 1.5	1.2% 1.9
Itemized	4.2	4.0
$10,000 - $15,000	2.0	2.8
$15,000 - $50,000	2.2	4.1
$50,000 & Over	11.0	32.3
BUSINESS		
Under $10,000	3.6	5.1
$10,000 - $30,000	2.5	1.0
$30,000 & Over	10.2	15.8
TOTAL	2.1	4.1
CORPORATIONS (Asset Class)		
NO BALANCE SHEET	3.5	6.8
UNDER $50,000	3.4	5.7
$50,000 - $100,000	6.4	12.6
$100,000 - $250,000	7.5	8.9
$250,000 - $500,000	14.3	8.3
$500,000 - $1,000,000	19.2	19.6
$1,000,000 - $5,000,000	36.9	30.2
$5,000,000 - $10,000,000	49.2	24.1
$10,000,000 - $50,000,000	45.3	31.0
$50,000,000 - $100,000,000	73.8	100.0
$100,000,000 & Over	100.0	100.0
TOTAL	9.4	10.6
FIDUCIARY	0.9	1.0
ESTATE (Audit Class)		
UNDER $300,000	14.0	8.1
$300,000 & Over	52.8	23.0
TOTAL	19.2	10.1
GIFT	3.8	3.9
TOTAL INCOME, ESTATE AND GIFT	2.4 [3]	4.2
EXCISE	11.1	11.1
EMPLOYMENT	0.6	1.6
TOTAL PROGRAM	2.1	3.7
PARTNERSHIPS (memo)	1.7	2.0

1/ Current services (not optimized)
2/ Plan optimized for equal marginal yield-to-cost ratios, except for classes
   with low VCL. Coverages include Service Centers DIFF CORR. examinations.
   These coverages are predicated on the realization of anticipated
   productivity improvements from the implementation of TAS.
3/ Includes DIF Corr. examinations at Service Centers not shown in
   individual class coverages.

should have been reported will be reduced by a specified amount. (Over-payments are counted as negative addends in the "gap".) Once these resources are allocated, the remainder of the manpower available is distributed among classes to maximize net yield. The methodology for this maximization will be discussed in the next section. Table 6 shows the "target" voluntary compliance levels of each class.

The tax-gap for each audit class is estimated based on the total dis-closable increases and decreases on all returns filed (projected from the TCMP Sample findings as if all returns filed were audited). To obtain the dollar amount of the tax gap for some future year, the tax gap of the TCMP year is adjusted to reflect the projected increase in the reporting population. Thus to project how much resources will be allocated to low complying classes in 1983, the following computation is made:

$$\text{1983 Tax Gap} = \frac{\text{\# of Returns filed in 1982}}{\text{\# of Returns filed in 1974}} \times \text{TG}$$

where TG = Tax Gap calculated for returns filed in 1974 (based on TCMP Sample of 1973 Returns)

Projections of 1981 returns filed in 1982 are used in the 1983 Tax Gap calculation because audits usually take place in the year following that in which returns are filed. For classes in which the voluntary com-pliance level is less than 90%, additional resources are allocated based on the following rules:

If Voluntary Compliance is Between	Amount of Gap to Recover	Resulting Compliance
85-90%	5%	86-91%
80-85%	10%	82-86%
75-80%	15%	79-82%
70-75%	20%	74-76%

These allocations are constrained by the requirement that ratios of marginal yield to marginal cost do not fall below 1 - 1.

It is not possible to assess the impact that the implementation of this plan will have on coverage levels, no-change rates, manpower alloca-tions, or compliance, because at the present time the necessary data are not available. We can, therefore, evaluate this approach only on a conceptual basis.

TABLE 6

DATA USED FOR DEVELOPING COMPLIANCE COVERAGE CURVES FOR
1974 AND SUBSEQUENT YEAR LONG RANGE PLANS
INDIVIDUAL RETURNS *
PHASE 3 TCMP POINTS

AUDIT CYCLE	1 (1963)		2 (1965)		3 (1969)		4 (1971)	
CLASS	Audit Coverage	Voluntary Compliance Level	Audit Coverage	Voluntary Compliance Level	Audit Coverage	Voluntary Compliance Level	Audit Coverage	Voluntary Compliance Level
AGI (000)								
NBU-10 standard	3.4	95.6	2.3%	95.4%	1.3%	95.2%	-	-
itemized	6.2	92.5	5.4	91.9	2.2	88.5	2.5	86.0
10-50	8.6	96.1	7.4	96.6	3.3	96.1	1.6	95.9
50 & over	28.5	95.6	27.2	95.8	15.4	94.1	-	-
BU-10	2.7	78.0	3.2	78.0	2.6	68.7	2.2	63.5
10-30	6.8	89.4	6.6	90.7	4.0	87.8	-	-
30 & over	17.8	91.5	19.1	93.3	12.9	91.2	-	-

CORPORATION RETURNS
PHASE 4. CYCLE 1 TCMP
(Returns filed in 1969)

Asset Class	COV.	VCL
Under $50,000	2.9%	59.9%
$50,000-$100,000	5.6	73.8
$100,000-$250,000	9.7	84.3
$250,000-$500,000	16.8	83.4
$500,000-$1,000,000	23.9	87.9

ESTATE TAX RETURNS FILED IN 1971

Gross Estate	COV.	VCL
Under $300,000	14.1%	81.5%
$300,000 & over	56.0%	86.9%

* The coverage compliance relationship for a particular year's compliance is related
to prior year examination coverage.

On face value, this program of action seems to be conservative, and workable as an interim measure. It places resources in low complying classes on the basis of how much direct compliance can be obtained, and makes no assumptions about what effects the application of this scheme may have indirectly on future compliance. The availability of better data and behavioral theories in the future are, however, required to provide a solid basis for reliance on this approach, improvements in the settings of its parameters, or (perhaps most likely) the directions for a better long-term approach.

The only Audit Planning goal which does not seem to be directly addressed by this plan is that of promoting a high degree of <u>voluntary</u> compliance. What is accomplished is the assurance that <u>after-audit</u> compliance is at a satisfactory level. Since little information presently exists which would allow the formulation of any program promising the promotion of high voluntary compliance, this strategy seems reasonable. However, it should be considered an interim improvisation to be replaced as soon as sufficient information exists by a strategy that <u>does</u> directly address the voluntary compliance issue.

The discussion in this section has, so far, addressed the overall strategy of the Audit Plan. We now discuss the procedures to be used in implementing this plan.

5.1 Implementation of Plan

The first step of the planning process consists of determining the maximum attainable level of proficient manpower. Account must be taken of the experienced manpower presently available, and of estimated attrition. Then plans must be made to recruit personnel with the proper education or experience and to train them in audit techniques. The number of new auditors is limited by (1) availability of qualified candidates, (2) budget, and (3) capacity of IRS training facilities. With the maximum level of trained available manpower determined, audit personnel are first assigned to the low complying classes. The definition of the Amount of Gap to Recover insures that not all the availablemanpower is allocated to the low complying classes. The manpower remaining is then allocated among all classes on the basis of maximization of net-yield. To determine the optimal mix among classes, coverage-to-yield curves for each audit class must be constructed.

Returns are ranked by their DIF scores. Thus, given an efficient DIF, the higher the score on a return, the greater the likelihood that the return will offer a potential tax change. It is therefore anticipated that as the number of returns examined within a class increases, the average yield per return decreases. The TCMP sample returns are used to determine this functional relationship as described next.

IRS divides DIF scores into 25-unit intervals in descending order of DIF scores. Then the audit coverages relating to these intervals and the resulting average yields for the intervals are computed. These data are

fed into a least-squares computer program to determine the best param-
meter fit of coverage-yield curves of pre-assigned form. For individual
returns the curve has the form $Y = a + b(\log X) + c(\log X)^2 + d(\log X)^3$,
whereas for corporate returns the curve has the form $\log Y = a + b(\log X) + c(\log X)^2$; here Y = average yield in dollars and X = coverage (%).
These curves are then adjusted to include an assumption of 3.3% pro-
ductivity growth per year (direct revenue/return).*

In 1976, IRS asked the National Bureau of Standards' Statistical Engi-
neering Laboratory to explore the possibility that other functional
forms might better fit the data. Two classes, the medium business class
and the low nonbusiness class (standard deduction) were examined.
The report, which examined two other models, stated that "An analysis of
the residuals showed both models to be statistically unsatisfactory for
both sets of data." The report states further that "obvious disconti-
nuities in the data make fitting any model difficult." Figure 1
shows a plot of the data for the medium business class and one of the
curves which were fitted to those data. These data lead to an un-
acceptable conclusion, namely, that if 4% coverage were applied, the
yield per return would be significantly less than if 5% coverage were
applied. This discontinuity occurs within the most critical portion of
the curve, i.e., in the feasible region of coverage. Additional plots
in Appendix C show similar discontinuities for the high nonbusiness,
medium farm business, and high nonfarm business classes. These data
further confirm that the current discriminant function analysis pro-
cedure is not ordering the returns efficiently.

IRS has decided to continue with its past formulation of yield/cost
curves, because within the region of greatest applicability these
curves fit the data at least as well as the new curves suggested by NBS.
We suggest that if, at the conclusion of the audit-class definition
study, IRS implements new Audit Class definitions or new DIF formulae,
then IRS should reevaluate the functional forms used to model the
coverage-yield relationship. It should be noted that TCMP data were not
required nor developed for corporations with assets from $1 million to
$50 million because of the high level of coverage of this class;
coverage-yield curves for these classes are calculated from data ob-
tained through examinations during previous fiscal years.

Once estimates of yield (as a function of coverage) are calculated, the
next step in the process requires the examination of audit costs. For
each audit class, a cost per return examined is calculated. This annual
estimate includes adjustments for fringe benefits, mix of technical
personnel, and types of examinations (field, office, correspondence) as
well as a 25% add-on in cost per return to reflect the costs of Appel-
late, Technical, Intelligence, Collection, data processing and other
support functions.

---

*The 3.3% production level assumed in 1975 was reduced to a 2.0%
production level in 1976, and a 0% production level in the 1977 LRP.

MEDIUM BUSINESS CLASS

YIELD PER RETURN
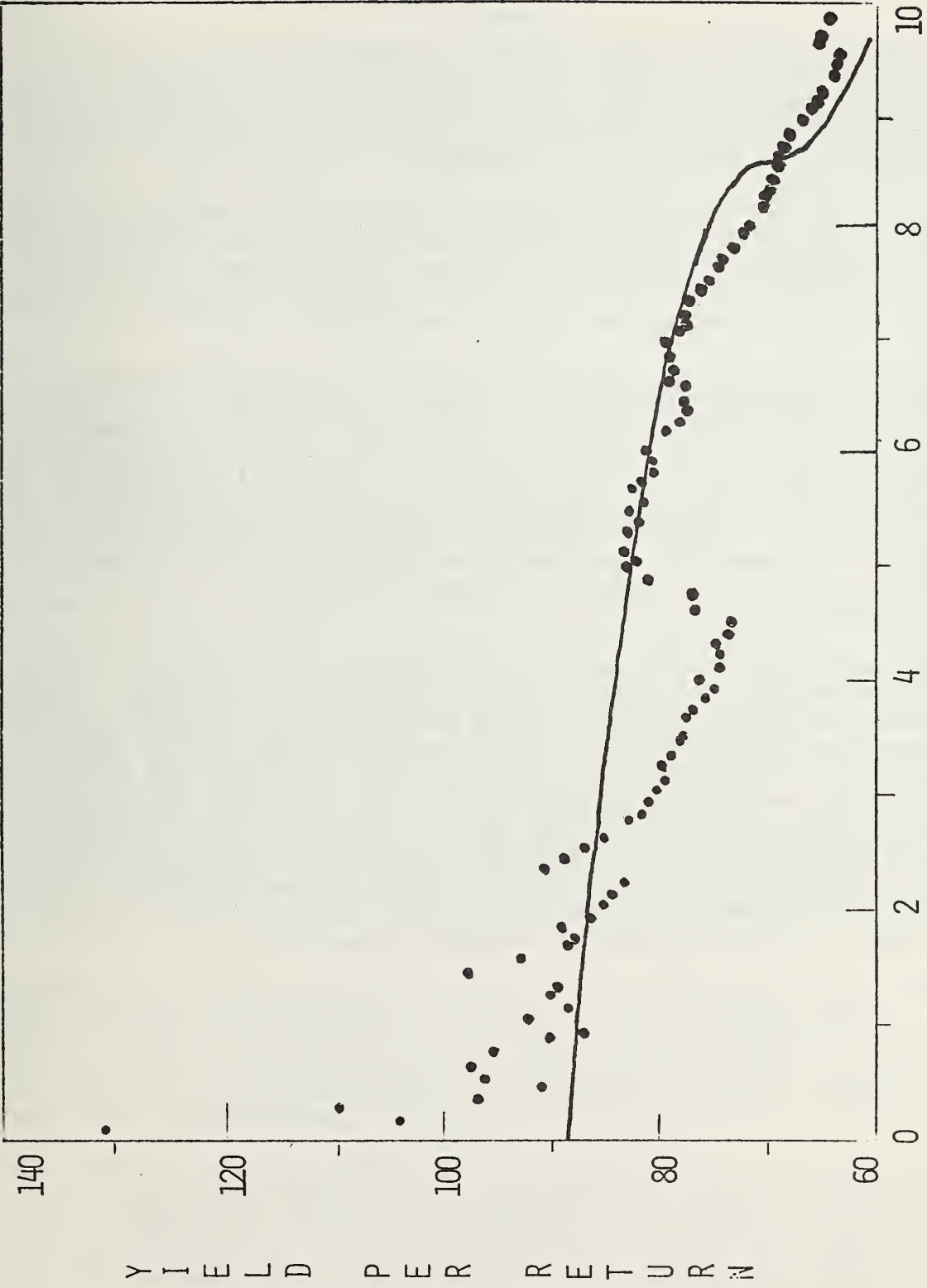
COVERAGE LEVEL
FIGURE 1

As implied above, the cost estimation procedure employs an assumption that cost per return is not a function of the level of coverage within an audit class. IRS is initiating a study to test the validity of this assumption, and if the assumption is found to be invalid, will develop the appropriate functional relationships.

Besides the above-mentioned study, we believe a study should be made of the effects of including in the cost calculations a cost to taxpayers who are audited and found to have little or no tax change. This would permit incorporation of the idea of "harrassment" into the audit strategy, thereby enhancing the "equity" goal of IRS.

Once yield/cost curves have been calculated, IRS uses an iterative process to determine the optimal allocation of resources. It begins with tentative coverage levels for the available manpower and tests the related marginal yield-to-marginal cost ratios. Shifts in manpower are made in small amounts between classes until equality in these ratios has been achieved. Computer programs have been developed to assist in these calculations. Although this method is basically heuristic in its approach, it seems to be performing satisfactorily and we believe that efforts at improvement are not now indicated in the face of other priorities.

The preceding discussion has briefly summarized the methodology which IRS uses to distribute manpower resources among the individual, corporate and estate tax returns. No data presently exist to allow similar planning of manpower-resources in the fiduciary, gift, excise and employment tax areas. Fiduciary returns were sampled for the first time in the TCMP which began in the summer of 1975, and these will supply data necessary to allow this tax area to be evaluated on a comparable basis in the future. Since total yield/total cost data do exist for these other tax areas (those of gift, excise and employment returns), manpower is allocated to them based on past operating results and evaluation of trends projected from past operations.

6. Summary

It has been our purpose in this report to present an overview of the IRS Audit Division's current methodology and to assess its Long Range Plan. To this end, we have discussed in detail the methods IRS currently uses to select returns for audit and some of the differences in the future selection methods proposed in the 1977 long range plan. We have examined the concept of voluntary compliance and what levels IRS has been able to achieve. This has led to a consideration of random (TCMP) audits as well as those mandated by DIF scores. We have suggested several areas of investigation concerning the relation between these audits and voluntary compliance.

We can characterize the Audit Division's response to the specific need to filter likely candidates for audit from the population of tax returns, and to general problems of planning resource allocation, as innovative, open minded and technically sound. We found, however, that documentation was neither easily available nor well catalogued, and that we had to rely too heavily on personal interviews. Departure of key personnel would make it near-impossible to ascertain the origins of, or rationales for, a number of procedures.

The most frequent complaint in surveys of Federal modelling activities is that while much description of supporting analysis and theory is made available through reports and journal articles, user manuals and computer program listings are hard to find. Ironically,in the IRS a complementary situation exists: the end result of studies is an operational procedure to be used internally, which therefore must be (and is) explicitly described so it can be put to use. However, documentation of the background theory and experiments supporting adoption of these procedures (rather than others) is very sketchy and vague. In either situation, the main reasons for the deficiency is clear--inadequate funding and priority to allow the documentation, and time pressures to complete the research and to implement the resulting operating procedures.

A transcript of work in progress reduces duplication of effort and time spent for orientation required by turnovers of the in-house project staff, or by participation of outside people in the R & D effort. Even in the absence of such circumstances, the transcript affords a review of the progress of experimentation for the analysts themselves--sometimes a second look discloses new insights.

It is not at all necessary for such a working document to be "polished" into suitability for publication, but it should be edited at least once with twin objectives of minimizing ambiguity and achieving conciseness in the exposition. In addition, such a document should be as self-contained as possible, so that a future reader need not be forced to hunt for background rationale.

In this report, notice has been taken of a number of methodological practices which, so far as we are able to see, are based on sound

principles (e.g., the general philosophy of stratifying returns into
audit classes and the allocation of resources based on a balanced
strategy). In addition, we have designated areas in which we believe
that research studies are likely to lead to significant improvements
(e.g., altering the audit class definitions, reformulating DIF formulae,
and understanding of taxpayer behavior). These suggestions are reca-
pitulated in tabular format at the end of the text.

The Audit Division's present Long Range Plan has evolved from years of
working experience and experimentation. It is, therefore, unlikely that
off-the-cuff suggestions for specific changes will bestow immediate
improvements; with no solid evidence to support such suggestions, we
refrain from making any. This evaluation has, however, highlighted areas
in which we believe there are plausible alternative approaches which may
in fact be preferable. Models or theories underlying these alternative
approaches often cannot be tested because of lack of data. If IRS is
committed to the goal of formulating an optimal or near-optimal audit
strategy for the long term, then the generation of such data is essen-
tial, and the in-depth research efforts to produce this information
should be initiated.

TABLE 7

SUMMARY OF RESEARCH TOPICS

1. Supplementary information input to the discriminant function analysis:

   a. Additional variables from tax forms (and possible broadening of tax form)
   b. Formalization of manual classification methods
   c. Tables of economic, socio-economic data

2. Modification of the DIF methodology:

   a. Distribution Free Methods (e.g. Kendall)
   b. Regression-like methods aimed at ranking tax returns.

3. Audit Class Redefinition Study:

   a. Alternative class boundary criteria
      e.g., Geographical, line items other than AGI, gross receipts, industry or occupational groups, primary source of income, etc.
   b. Separate (additional) DIF scores for most important tax schedules in 1040.
   c. Cluster Analysis Approach

4. Compliance related research:

   a. Surveys of taxpayers' perceptions
   b. Controlled studies to evaluate effects on voluntary compliance levels of media campaigns, educational materials, etc.
   c. Game theoretic analysis - see Appendix A
   d. Double hits (effect of repeated audits)
   e. Ripple effect (e.g., within an occupational group)

5. Implementation of the Audit Plan:

   a. Explore audit-yield relationship based on trial redefinitions of audit classes.
   b. Incorporate "harrassment" into model (e.g., add a "penalty" cost in calculating yield/cost). See Appendix B.
   c. Examine field implementation of plan.

REFERENCES

[1] Anderberg, Michael R., Cluster Analysis for Applications, 1, Academic Press, 1971.

[2] Dixon, W.J. and Jennrich, R.I., "Computer Graphical Analysis and Discrimination," Discriminant Analysis and Applications, Edited by T. Cacoullos, Academic Press, 1972, pp. 161-172.

[3] Duran, Benjamin and Odell, Patrick, Cluster Analysis, A Survey, Lecture Notes in Economics and Mathematical Sciences #100, Springer-Verlag, 1970.

[4] Kendall, M.G. "Discrimination and Classification," pp. 168-185 in Multivariate Analysis, edited by P.R. Krishnaiah, Academic Press, 1966.

[5] National Opinion Research Council, "Taxpayers' Attitudes Towards Enforcement and Cheating on Personal Income Tax Returns: September 1966," a report submitted to the IRS' Research Division as partial fulfillment of contract TIR-25197.

[6] Schwarz, Richard D. and Orleans, Sonya, "On Legal Sanctions," The University of Chicago Law Review, V 34, pp. 274-300.

APPENDIX A

GAME THEORY AND AUDIT:
A PRELIMINARY EXAMINATION OF A MATHEMATICAL TOOL
FOR ANALYSES OF AUDIT STRATEGY AND TAXPAYER COMPLIANCE

Game theory is one of several mathematical models used to study decision making. (Other such models include classical decision theory, utility theory and simulation, i.e., gaming.) Game theory is particularly applicable to situations in which there is a conflict among several parties. It will not be necessary for us to consider anything other than the simplest case, namely, a conflict between two parties. As we shall consider it, a game consists of two players, together with a set of rules of play and a scoring mechanism called a pay-off function. Both players are aware of the rules of the game and the pay-off function. Each player makes a decision - that is, he selects one course of action from among those available to him according to the rules. In the language of game theory, each course of action which does not violate the rules of the game is called a strategy. Of course, each player must choose his strategy without the knowledge of what the other player's strategy will be. The pay-off function, with the two chosen strategies, then determines what "payoff" each player receives.

There are many well-known examples of two-person games that fall within the broad outline given in the last paragraph. Perhaps the simplest of these is the children's game of matching pennies. Each player shows a coin with either the head face or the tail face visible. The rules of the game call for one player to win if both players' coins show the same face while the other player wins when the faces on the coins are different. Here each player must decide how often (i.e., with what probability) to play heads and how often to play tails. The pay-off is +1 for a player when he wins and -1 when he loses.

Clearly, each player in a game should decide in such a way as to maximize (in some sense) his pay-off. In fact, game theory assumes each player is a "rational person" who will do so. (This usage of "rational" comes from classical economics.) The philosophy adopted in game theory is, however, also extremely conservative, namely, each player plays so as to maximize the expectation that he is able to <u>guarantee</u> for himself, irrespective of the other player's decision. Thus, as each player plays in such a way as to guarantee a certain amount, he may have to forego the (non-guaranteed) possibility of larger payoffs. For each player, a strategy which maximizes the amount that he is able to guarantee for himself, is called an optimal strategy. A game is said to be solved when each player's optimal strategies (and the corresponding pay-offs) have been found. The Fundamental Theorem of Game Theory insures that a broad class of games do actually have solutions. The games described here are simple representatives of two-person games, and the existence of solutions for them is assured by the Fundamental Theorem.

In matching pennies, each player can guarantee an expectation of at least zero for himself by selecting heads and tails with equal probability, but in some random way (e.g., by tossing his coin in the air and letting it fall as it may). It can be shown that in this game no strategy exists for either player which will guarantee an expectation greater than zero. Thus, for each player, playing heads and tails each with probability 1/2 is an optimal strategy and the corresponding pay-off for each player is zero.

In a recent study [1], [2] conducted by the National Bureau of Standards, game theory was applied to "optimizing" the inspection strategy of state agencies concerned with regulating the accuracy of measuring devices used in retail trade. Three models were constructed reflecting different possible goals of the regulating agency. The general type of scenario examined in [1], [2] appears to be adaptable to the audit situation within IRS. One player in the game is IRS, whose decision (i.e., choice of strategy) consists of selecting which tax returns to audit, up to the limit of its resources. The other player is the taxpayers of the country. A. J. Goldman and D. R. Shier [3] have shown that such a set of taxpayers can be considered as a single player for the purposes of game-theoretic analysis. The second player's decision consists of selecting those tax returns on which to "cheat".*

In the Audit context, applying the game theoretic notions of "rational behavior" and "maximization of expectation" does not necessarily signify the presence of fraudulent behavior. Any entry in a tax return for which the taxpayer recognizes the possibility of a dispute by IRS (such as giving oneself the benefit of the doubt in interpreting tax regulations, conscious carelessness, "not bothering" to check tax regulations or deliberate nonfeasance) may represent a rational decision in the standard context of game theory.

---

*It was found convenient in [1] to use the ugly word "cheat" to denote any malfunction of a measuring device, either by design or by inadvertence. We maintain here the convention that "cheating" covers any situation in which the taxpayer makes any entry in his tax return for which he is conscious of the possibility that it might be challenged by IRS.

True error, or inadvertence without any overtones of violation, could be accommodated within game theoretic structures, but since its inclusion in a game would require some complications of probability analysis, we will for simplicity ignore its existence in the present discussion.

The rules of the game are the IRS regulations, including the penalties imposed when errors are discovered through a tax audit. The pay-off to the tax-payer is the profit that he can expect to receive (including a loss as a negative profit) by the occurrence of deficiencies in his tax return.

There are several choices for how the pay-off function to the IRS should be formulated to reflect that Service's "mission." The simplest one (and the first one considered in [1]) is for the taxpayer's gain to be IRS's loss. Thus, following this scenario, IRS's aim is to minimize the taxpayer's gain due to cheating. A second scenario (and the second one considered in [1]) calls for IRS to minimize the tax-payer's underlined cheating - with the rationale that taxpayers once forced to pay additional taxes, interest and penalties as a result of a tax audit are less likely than other taxpayers to cheat in the future. A third scenario treated in [1] leads to a game in which it is assumed that the taxpayer has become familiar with IRS's strategy before he chooses his own strategy - and that IRS knows this. Thus, IRS wishes to elect that strategy which, the taxpayer being aware of it, yields to the taxpayer the smallest gain due to cheating.

A reader who is frightened of mathematical notation may omit the material set off by quadruple asterisks **** without disrupting the flow of ideas, but at the risk of escaping knowledge of how game theory works in the current context.

We will number the tax returns from 1 to n (at the moment, no particular rank criterion for the tax returns is implied in this ordering). For the ith tax return, we define the following terms:

V_i - the dollar Value to the taxpayer of undetected cheating, i.e. the amount of change in tax liability which would take place were the tax return to be audited,

P_i - the amount (Penalty, etc.) assessed against the taxpayer when cheating is discovered on his tax return by audit,

c_i - the probability of cheating,

p_i - the probability that his return will be audited by IRS and any cheating discovered.

<div align="center">****</div>

Also, we denote IRS audit resources by m; that is, m is the largest number of tax returns that can be audited by IRS. If specific amounts of audit resources are preassigned to the audit classes (or to any other partition of the tax returns) then we denote by m_j the largest number of tax returns of the jth class that can be audited by IRS. The sum of the m_j's must be at most m.

Since c_i and p_i are probabilities, we must have

$$0 \leq c_i \leq 1 \quad \text{and} \quad 0 \leq p_i \leq 1.$$

The limitation on IRS's audit resources is expressed by the inequality

$$\sum_{i=1}^{n} p_i \leq m.$$

In other words, the expected number (in the technical sense) of returns to be audited cannot exceed the resource limits. Any set of numbers c_1, c_2, ..., c_n, subject to the restriction $0 \leq c_i \leq 1$ constitutes a strategy for the taxpayer. A strategy for IRS is any set of numbers p_1, p_2, ..., p_n, subject to the two conditions

$$0 \leq p_i \leq 1 \qquad \text{and} \qquad \sum_{i=1}^{n} p_i \leq m.$$

When the taxpayers and IRS have chosen strategies (call them c and p respectively), then the expected pay-off to the ith taxpayer (and consequent loss to the Treasury) is $(V_i - p_i P_i)c_i$. Consequently, the expected total loss to the Treasury is

$$L(p,c) = \sum_{i=1}^{n} (V_i - p_i P_i)c_i.$$

For each specific strategy that IRS adopts (call it $p^{\#}$), there is a maximum expected total loss to the Treasury (which depends on $p^{\#}$) that can occur, namely, the maximum value that $L(p^{\#}, c)$ takes on as c varies over all possible strategies for the taxpayers. Application of game theory in this context calls for finding $p^{\#}$ for which the maximum expected total loss is as small as possible.

We were told by IRS that the assessment against the taxpayer, when cheating is detected on his return by an audit, is the sum of (1) the

41

unpaid tax, (2) an interest charge, and (3) a penalty.* Since (2) and

(3) are fixed percentages of (1), it follows that the value of V_i/P_i is

the same for all detected cheating. Let us denote the value of that quotient

by q. Clearly $0 \leq q \leq 1$.

In order to apply the model described above to the IRS audit situation,

it is necessary to have numerical values for m, n, V_i and P_i. Equivalently,

it is necessary to have values for m, n, V_i and q. Each player will need

to know some (or all) of these values in order to construct his optimal

strategy. The values of m, n and q are known to IRS; n and q are (prob-

ably fairly accurately) perceived by the taxpayers. However, IRS

is not able to determine V_i very readily since, according to the definition

of V_i, its value depends on the outcome of an audit of the ith tax re-

turn. In order to estimate the value of V_i for each tax return, a mecha-

nism is needed which, like the DIF, would determine a gain potential for

a return from its entries. Only after an audit takes place can IRS de-

termine the accuracy of this estimate. Since the definition of cheating

that we have used requires that the taxpayer be aware of the possibility

of IRS disputing the entries on his tax return, the taxpayer has more

information than IRS about V_i before an audit takes place.

Audit coverage typically varies widely over audit classes; in recent

years, aggregate coverage has averaged about 2%, i.e. m = .02 times n.

The results of [1] imply that (in the IRS audit context), whenever q

(i.e., V_i/P_i) is greater than m/n then the optimal strategy for the

taxpayer is to cheat in virtually all cases. In order for the taxpayer's

_____

*In the game model, the penalty which we have called P_i consists of the
 sum of (1), (2) and (3).

42

optimal strategy to call on him not to cheat, it is necessary that
either m be increased or that q be decreased (by increasing the penalty,
P_i), or both, until q is less than m/n.

Here is a sample "play" of a realization of the prototype game, whose
parameter values are derived from IRS data. One player is "the auditor".
The other player is "the taxpayer" with 1000 tax returns (i.e. n = 1000).
The values of the 1000 V_i's ("taxpayer's potential gain" part of the pay-
off function) were chosen to follow fairly closely the national percentage
distribution of actual "tax change" amounts of one audit class: - "Under
$10,000 AGI, individual, itemized" in the Phase III TCMP survey. (See
histogram, Fig. A1.) The values of P_i were set by multiplying by 1.1 the
value of the corresponding V_i (i.e. penalty + interest = 10% of de-
ficiency; q = 10/11). The value of m was set at 40, or 4% of n (the
FY 1981 audit coverage level designated for the specimen audit class
under the 1976 long-range audit plan). The numerical specification of
the game, optimal strategies of the players, and the value of the
games are listed in Table A1.

A solution* for this game calls for the auditor to distribute his re-
sources uniformly over the $m' = m/q$ (= 40 x 11/10 or 44) returns with
largest values of V_i, that is, to audit each of these returns with relative
frequency q = 10/11 or probability p_i of approximately .909, and not
to audit the remaining $n-m'$ (=956) returns. The other player cheats,
that is, the taxpayer understates his liability by V_i, for the 956 returns
which will not be audited under the auditor's solution strategy. With

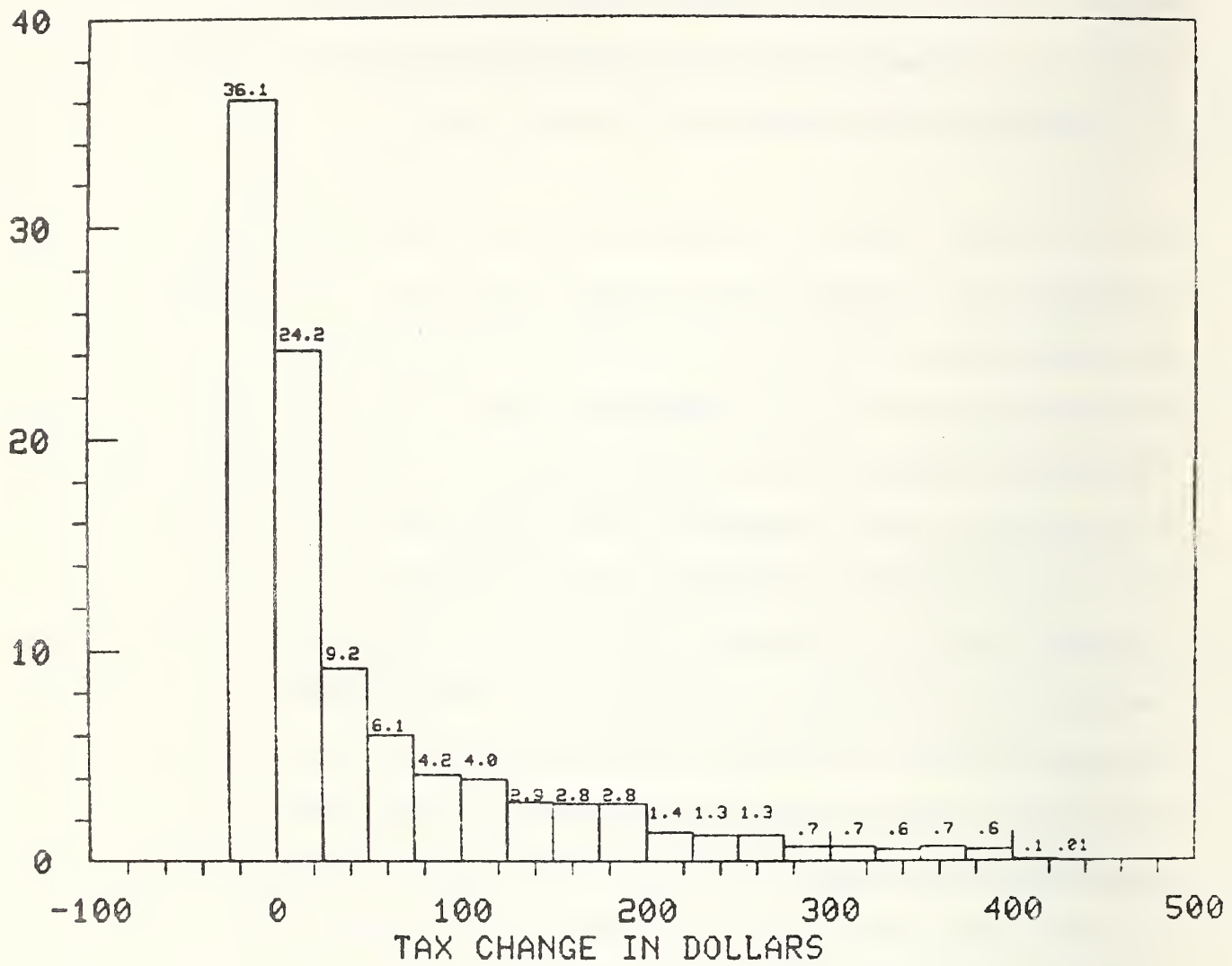―――――――――――
*We have not determined that it is unique.

43

FIGURE Al

Percent Distribution of Returns According to Tax Change Audit Class:
Under $10,000 AGI-Itemized-Non Business Derived from TCMP Sample 1966

(The first tax change interval, marked between -25 and 0, represents
the zero-change returns. The total on the chart is less than 100.0%
because of unplotted very small percent values for large tax changes.
The 5 largest tax changes actually recorded in the sample were $67,964,
$25,897, $25,291, $11,696 and $11,300.)

44

GAME PARAMETERS   SOLUTION VALUES

i	$V_i(\$)$	$P_i(\$)$	c_i	p_i
Tax Return Index Number	Potential Gain to Taxpayer	Potential Cost to Taxpayer (V_i/q)	Relative Frequency (or Probability) of "CHEAT" (Taxpayer Decision)	Relative Frequency (or Probability) of AUDIT (Auditor Decision)
1	10000	11000	350/10000 (.035)	10/11 (.909)
2	7000	7700	350/7000 (.050)	"
3	5000	5500	350/5000 (.070)	"
4	3500	3850	350/3500 (.100)	"
5	2500	2750	350/2500 (.140)	"
6	1500	1650	350/1500 (.233)	"
7	1000	1100	350/1000 (.350)	"
8	900	990	350/900 (.389)	"
9-10	800	880	350/800 (.438)	"
11-13	700	770	350/700 (.500)	"
14-17	600	660	350/600 (.583)	"
18-23	500	550	350/500 (.700)	"
24-32	450	495	350/450 (.778)	"
33-44	400	440	350/400 (.875)	10/11 (.909)
45-54	350	385	1/1 (1.0)	0 (0.0)
55-70	300	330	"	0 "
71-79	275	302.5	"	0 "
80-91	250	275	"	0 "
92-105	225	247.5	"	0 "
106-123	200	220	"	0 "
124-143	175	192.5	"	0 "
144-162	150	165	"	0 "
163-202	125	137.5	"	0 "
203-244	100	110	"	0 "
245-305	75	82.5	"	0 "
306-397	50	55	"	0 "
398-639	25	27.5	"	0 "
640-1000	0	0		0

$n = 1000$ $m = 40$ $q = 10/11$ $L = \$52,800$

Total potential Treasury loss = \$101,400

TABLE A1:   NUMERICAL VALUES FOR SAMPLE GAME

reference to the game parameters, this means that $c_i = 1$ for these

returns. For each of the other 44 returns, the taxpayer assigns a

frequency of cheating equal to the ratio between the 45th (m' + 1st)

largest V_i (or P_i) and the V_i (respectively, P_i) of the return in ques-

tion, yielding c_i's which range from about .9 down to about .04. The

value of the game, L, "loss to the treasury" under the solution strate-

gies, is the sum of the V_i's for the 954 returns which are destined to

be ignored by the auditor, or $52,800. Each of those returns contributes

V_i to the sum

$$\sum_{i=1}^{n} (V_i - p_i P_i) c_i,$$

being of the form $(V_i - 0 \cdot P_i) \cdot 1$. For each of the other m' (or 44)

returns the expected (average) contribution is zero because under the

auditor's optimal strategy the term in parentheses becomes $V_i - q \cdot P_i =$

$V_i - q \cdot \dfrac{V_i}{q} = 0.$

The total potential taxpayer gain (the sum of all the V_i) is $101,400,

and so the game value represents abatement of 47.9% of the potential de-

ficiency, by the optimal audit strategy.

If the "penalty rate" in this game were doubled, i.e. $P_i = 1.2\ V_i$

instead of $P_i = 1.1\ V_i$, then, given the coverage rate of 4%, we would

have m' = 48 instead of m' = 44; the number of unrestricted cheats

would drop by 4 and the remaining cheat frequencies reduced slightly;

the value, L, of the game would be reduced by $1400 and the expected

deficiency abatement would rise to slightly above 49%.

To restrict the propensity to cheat to those returns with V_i no larger than $100 would require a value of q of about 1/5, or a penalty + interest rate of about 400% of deficiency, producing $19,452 or 80.8% abatement. A target of "no cheating", but with a "no change" criterion set at "$25 or less" would require about 900% penalty + interest (L = $6050; 94% abatement), while reducing the game value to literal zero would require a penalty rate of 1500%.

In interpreting this example, bear in mind that the values of V_i were determined by invoking recorded <u>realized</u> levels of post-audit tax deficiency to stand for <u>potential</u> deficiency levels, and that the rules of the game imply that both players have full knowledge of the values of n, m and all the V_i and P_i. Notice, thus, that the solution strategy for the taxpayer assigns the smallest probability of cheating to the returns with the largest values of V_i, meaning that these will pose the least temptation for the taxpayer. This possibly startling idea becomes intuitively reasonable when one realizes that according to the auditor's solution strategy these are the returns most likely to be audited. The fact remains, however, that according to the data base all of those high returns <u>did</u> "cheat". These caveats aside, the sample game solution suggests one possibly useful idea that may have been overlooked heretofore: that the solution strategy for the auditor tells him <u>not</u> to audit all of the m "highest ranking" returns but to extend his attention to a slightly wider group and audit them all "almost" all the time. The "taxpayer's" response in reducing his cheat frequencies implies that under some actual circumstances IRS might obtain a "bigger bang for the buck" in

terms of inducing compliance by spreading the audit resources slightly thinner but further at the high end of the DIF. This completes discussion of the example.

<div align="center">****</div>

Fitting the above analysis framework to the audit situation within IRS will not be straightforward. For example, the conservative philosophy which underlies game theory may not be an accurate reflection of the thinking of the American taxpayer and may be too conservative a strategy for IRS as well. Also, it is clear that more is involved in the proper evaluation for each taxpayer of V_i (gain from understating tax liability) and P_i (loss through penalty) than the amount of money at stake. If there were no considerations other than monies paid in penalties (P_i) or underpayments of taxes (V_i) and if the players were fully aware of the rules of the game* (in particular, if the taxpayers were aware of the amount of audit resources now available to IRS), then an optimal strategy for the taxpayer today would call for virtually universal cheating. Why does this not, in fact, take place? Clearly the game model in the simple version described above does not adequately mirror the income tax situation as it currently exists. We will suggest several possible ways of improving the model, that is, making it reflect more accurately the behavior of the taxpayer. Further study, however, will then be needed to determine whether the new models (1) do indeed reflect more accurately the behavior of the taxpayer, and (2) are still mathematically tractable.

---

*It can be anticipated that, with the Freedom of Information Act in effect, taxpayers will soon become more aware of the true rules of the game.

One drawback of the model described above is that a single value of q, the same for all taxpayers, does not exist. There are many often intangible, aspects of V_i and P_i other than the numbers of dollars involved. We present here several other possible constituents of the "true" values of V_i and P_i:

$$V_i$$

1) Satisfaction from gambling and risk-taking.

2) One-upmanship (i.e., keeping up with one's friends and neighbors).

3) Evidence of one's cleverness and sophistication (i.e., "Beating the System").

4) "Revenge" on the government for following policies unpopular with the taxpayer.

5) Benefit of cash available now compared to the future possibility of loss.

$$P_i$$

1) Financial losses in addition to the assessment P_i (e.g. legal fees, time lost from work, etc.).

2) Loss of time in preparing for an audit and in being audited.

3) Conscience.

4) A generalized distaste for gambling and risk-taking.

5) Fear of exposure and publicity.

6) Fear of prosecution (even if the case should be decided in the taxpayer's favor).

7) Fear of being convicted of tax evasion (criminal record, prison, etc.).

8) Belief that the government is omnipresent and omnipotent (i.e., you can't fight City Hall).

We believe that these (and possibly additional) considerations play a very important role in the way that the taxpayer perceives and plays the game. Of course we have introduced a serious problem in requiring realistic values for both V_i and P_i. The distribution of pertinent attitudes and psychological traits among taxpayers must somehow be reflected in efforts to assign reasonable dollar equivalents to each of the items listed above. Clearly these values will differ from one taxpayer to the next, and a realistic model must allow for this. Research along these lines has already begun [4] and promises to continue in the future [5].

The problems and difficulties which we have just described indicate that the simple game-theoretic model presented earlier is not by itself an adequate tool with which to study the audit problem as related to the individual taxpayer. In particular, the assumption made in our game that the taxpayer's loss is the Treasury's gain (and visa versa) is not really valid. Game models which allow for differing gains among players (called non-zero sum games) have been studied extensively and this (somewhat more complicated) model might be more representative of the audit situation.

The non-zero sum model mentioned above as well as the first model previously described (and treated in [1] and [2]) call for the taxpayer to decide whether or not to cheat a specified amount, V_i. A second way in which the game described above does not accurately model the income tax situation is that no provisions exist in the model for the taxpayer to decide how much to cheat. A game which mirrors the IRS audit scenario more closely would permit the taxpayer to cheat any amount he wishes (including negative amounts), that is, the amount that a taxpayer will cheat is described by a probability function and, at least in theory, many different amounts of cheating are possible. Of course the taxpayer understands that the larger the amount that he decides to cheat, the greater may be the probability of audit. Incorporating this additional variable into the taxpayer's strategy would significantly alter the rules of the game and may require substantial further analysis in order to obtain a solution. For example, it is not clear whether the set of taxpayers could still be considered as a single player for the purpose of game-theoretic analysis. As before, IRS requires a mechanism (such as DIF) with which to estimate the tax change potential of a tax return, and an audit process to verify that the estimates given by this mechanism are reliable. Also, IRS must formulate a strategy, based on the rules of the game and the estimate of the taxpayer's strategy, with which to maximize its pay-off. Presumably, this strategy would require that the probability of audit for a particular tax return reflect the amount of cheating likely to have been selected in the taxpayer's strategy. Thus it is important that the tax-change potential of a return,

as estimated by IRS, correlate strongly with the amount of cheating in the taxpayer's strategy. A more accurate model of the income tax administration process would result from considering the probability that an entry on the tax return which the taxpayer knows is subject to dispute will eventually (either by the auditor or through litigation) be upheld.

More substantial changes in the game model may be needed in order to properly describe the income tax situation. It may be necessary to introduce new formulations of the rules such as bluffing, games with imperfect information, cooperative games, etc. These sophisticated models have been treated extensively in the mathematical literature and can be used to more accurately reflect the current income tax picture.

It is unlikely that any mechanical formulation can furnish entirely reliable forecasts of taxpayer compliance. We believe, however, that game theoretic analysis, which focuses on the divergence of objective which separates the taxpayer from IRS--or if you will, the taxpayer as economic creature from the taxpayer as citizen--is the formal technique which offers the greatest independent hope of profitable application, in addition to furnishing a yardstick for measuring representation of simply motivated "extreme cases" of taxpayers' behavior, in conjunction with other models.

REFERENCES

[1] A. J. Goldman and Pearl, M. H., "The Dependence of Inspection-
     System Performance on Levels of Penalties and Inspection Resources"
     Journal of Research of the National Bureau of Standards-B. Mathe-
     matical Sciences, 80B, 1976, pp. 189-236.

[2] M. H. Pearl and Goldman, A. J. "Policing the Market Place"
     Mathematics Magazine V50, No. 4, 1977, pp 179-189.

[3] A. J. Goldman and Shier, D. R., "Player Aggregation in Noncoopera-
     tive Games", working paper, Center for Applied Mathematics, National
     Bureau of Standards, Washington, D.C. 20234.

[4] Schwarz, Richard D. and Orleans, Sonya, "On Legal Sanctions" The
     University of Chicago Law Review V34, pp 274-300.

[5] "A Study to Determine the Relative Impact of the Internal and
     External Factors Affecting Compliance with Individual Tax Laws".
     Request for Proposal IRS 78-35 January 10, 1978.

APPENDIX B

SOLUTIONS OF THE GAME-THEORETIC MODELS

This technical attachment presents (and proves the validity of) particular solutions in symbolic form for two of the three basic game-theoretic models alluded to in Appendix A.   The first of these games was described in some detail in that Appendix and generates the numerical example of its text; a second one was mentioned briefly as a variant. The notation of Appendix A is maintained. Model 1, in this treatment, is slightly more general than its version in the foregoing material: the simplifying assumption of a constant ration V_i/P_i is discarded.

Model 1

For this model we use a zero-sum game, with objective function

$$F(c,p) = \sum_{i=1}^{n} (V_i - P_i p_i) c_i$$

for the taxpayer and its negative as objective function for the auditor.

Define q_i, $i=1,2,\ldots,n$ by

$$q_i = V_i/P_i.$$

Then

$$F(c,p) = \sum_{i=1}^{n} P_i (q_i - p_i) c_i.$$

We assume that the returns are numbered so that

$$P_1 \geq P_2 \geq \cdots \geq P_n.$$

Then the set T of "tempting" returns, whose size is denoted $|T|$, is defined by

$$T = \{i \mid q_i > 1\} = \{i \mid V_i > P_i\}.$$

The complementary set (of "untempting" returns) is denoted $\overline{T}$.

Case 1. The defining relation for Case 1 of Model 1, which describes the inspection resources as being below a certain threshold, is

$$m < |T| + \sum \{q_i \mid i \in \overline{T}\} = \sum_{i=1}^{n} \min\{1, q_i\}.$$

Let k be the largest integer for which

$$\sum_{i=1}^{k} \min\{1, q_i\} \leq m.$$

Then $0 \leq k < n$. Set

$$\alpha = m - \sum_{i=1}^{k} \min\{1, q_i\}.$$

Then $0 \leq \alpha < \min\{1, q_{k+1}\}$.

Theorem: An optimal strategy for the auditor is given by

$$p_i^{\,o} = \min\{1, q_i\} \qquad i \leq k$$

$$p_{k+1}^{\,o} = \alpha$$

$$p_i^{\,o} = 0 \qquad\qquad i > k+1.$$

An optimal strategy for the taxpayer is

$$c_i^O = 1 \qquad\qquad i \leq k, \, i \varepsilon T,$$

$$c_i^O = 1 \qquad\qquad i > k,$$

$$c_i^O = P_{k+1}/P_i \qquad\qquad i \leq k, \, i \varepsilon \overline{T}.$$

The resultant payoff is

$$F^O = F^O(c^O, p^O) = \sum_{\substack{i \varepsilon T \\ i \leq k}} V_i + \sum_{i > k} V_i - \sum_{\substack{i \varepsilon T \\ i \leq k}} P_i - P_{k+1}\alpha.$$

Proof: First we wish to show that

$$F(c, p^O) \leq F(c^O, p^O) \qquad\qquad \text{for all } c.$$

For any c,

$$F(c, p^O) = \sum_{\substack{i \varepsilon T \\ i \leq k}} (V_i - P_i)c_i + (V_{k+1} - P_{k+1}\alpha)c_{k+1} + \sum_{i > k+1} V_i c_i$$

since, for $i \varepsilon \overline{T}$, $i \leq k$, we have

$$V_i - P_i p_i^O = V_i - P_i q_i = V_i - P_i(V_i/P_i) = 0.$$

Because all $c_i \leq 1$, $V_i - P_i \geq 0$ for $i \varepsilon T$, and $\alpha < q_{k+1}$, it follows that

$$F(c, p^O) \leq \sum_{\substack{i \varepsilon T \\ i \leq k}} (V_i - P_i) + (V_{k+1} - P_{k+1}\alpha) + \sum_{i > k+1} V_i$$

$$= \sum_{\substack{i \varepsilon T \\ i \leq k}} V_i + \sum_{i > k} V_i - \sum_{\substack{i \varepsilon T \\ i \leq k}} P_i - P_{k+1}\alpha = F(c^O, p^O)$$

for all c.

Next we wish to show that

$$F(c^O, p^O) \leq F(c^O, p) \qquad\qquad \text{for all } p.$$

For any p,

$$F(c^O, p) = \sum_{\substack{i \varepsilon T \\ i \leq k}} (V_i - P_i p_i) + \sum_{i > k} (V_i - P_i p_i)$$

$$+ \sum_{\substack{i \varepsilon \overline{T} \\ i \leq k}} (V_i - P_i p_i) P_{k+1}/P_i$$

56

$$= \sum_{\substack{i \in T \\ i \leq k}} V_i + \sum_{i > k} V_i + P_{k+1} \sum_{\substack{i \in \bar{T} \\ i \leq k}} q_i - \sum_{\substack{i \in T \\ i \leq k}} P_i p_i$$

$$- \sum_{i > k} P_i p_i - P_{k+1} \sum_{\substack{i \in \bar{T} \\ i \leq k}} p_i$$

$$= \sum_{\substack{i \in T \\ i \leq k}} V_i + \sum_{i > k} V_i + P_{k+1} \sum_{\substack{i \in \bar{T} \\ i \leq k}} q_i - \sum_{\substack{i \in T \\ i \leq k}} P_i p_i$$

$$- \sum_{i > k+1} P_i p_i - P_{k+1} \left(\sum_{\substack{i \in \bar{T} \\ i \leq k}} p_i + p_{k+1} \right).$$

Since $\sum_{i=1}^{n} p_i \leq m$, it follows that

$$F(c^o, p) \geq \sum_{\substack{i \in T \\ i \leq k}} V_i + \sum_{i > k} V_i + P_{k+1} \sum_{\substack{i \in \bar{T} \\ i \leq k}} q_i - \sum_{\substack{i \in T \\ i \leq k}} P_i p_i$$

$$- \sum_{i > k+1} P_i p_i - P_{k+1} \left(m - \sum_{\substack{i \in \bar{T} \\ i \leq k}} p_i - \sum_{i > k+1} p_i \right)$$

$$= \sum_{\substack{i \in T \\ i \leq k}} V_i + \sum_{i > k} V_i + P_{k+1} \sum_{\substack{i \in \bar{T} \\ i \leq k}} q_i$$

$$- \sum_{\substack{i \in T \\ i \leq k}} (P_i - P_{k+1}) p_i + \sum_{i > k+1} (P_{k+1} - P_i) p_i - P_{k+1} m.$$

The first three terms of the sum do not depend on p. Since the P_i's are in descending order of magnitude,

$$F(c^o, p) \geq \sum_{\substack{i \in T \\ i \leq k}} V_i + \sum_{i > k} V_i + P_{k+1} \sum_{\substack{i \in \bar{T} \\ i \leq k}} q_i - \sum_{\substack{i \in T \\ i \leq k}} P_i \cdot 1$$

$$- P_{k+1} \left(m - \sum_{\substack{i \in T \\ i \leq k}} 1 \right)$$

$$= \sum_{\substack{i \epsilon T \\ i \leq k}} V_i + \sum_{i > k} V_i - \sum_{\substack{i \epsilon T \\ i \leq k}} P_i - P_{k+1} \left(m - \sum_{i=1}^{k} \min\{1, q_i\} \right)$$

$$= \sum_{\substack{i \epsilon T \\ i \leq k}} V_i + \sum_{i > k} V_i - \sum_{\substack{i \epsilon T \\ i \leq k}} P_i - P_{k+1} \alpha = F(c^o, p^o)$$

for all permissible strategies $\bar{p}$.

It now follows that (c^o, p^o, F^o) is a solution of the game. This concludes the discussion of Model 1, Case 1.

<u>Case 2</u>. The defining relation for Case 2 of Model 1 is

$$m \geq |T| + \Sigma\{q_i \,|\, i \epsilon \overline{T}\} = \sum_{i=1}^{n} \min\{1, q_i\}.$$

We also define

$$V_T = \Sigma\{V_i \,|\, i \epsilon T\}.$$

<u>Theorem</u>: Any strategy p^o for the auditor which satisfies the conditions

$$p_i^{\ o} = 1 \qquad\qquad i \epsilon T$$

$$p_i^{\ o} \geq q_i \qquad\qquad i \epsilon \overline{T}$$

is an optimal strategy. An optimal strategy for the taxpayer is given by

$$c_i^{\ o} = 1 \qquad\qquad i \epsilon T$$

$$c_i^{\ o} = 0 \qquad\qquad i \epsilon \overline{T}.$$

The resultant payoff is

$$F(c^o, p^o) = V_T - \sum_{i \epsilon T} P_i \ .$$

Proof: First we show that

$$F(c, p^o) \leq F(c^o, p^o) \qquad\qquad \text{for all } c.$$

$$F(c, p^o) = \sum_{i=1}^{n} (V_i - P_i p_i^{\ o}) c_i$$

$$= \sum_{i \epsilon T} (V_i - P_i) c_i + \sum_{i \epsilon \overline{T}} (V_i - P_i p_i^{\ o}) c_i$$

$$\leq \sum_{i \epsilon T} (V_i - P_i) = V_T - \sum_{i \epsilon T} P_i$$

since, for $i\varepsilon T$, we have $V_i - P_i > 0$ and, for $i\varepsilon \overline{T}$, we have $V_i - P_i p_i^{\,o} \leq 0$.

It is also clear that

$$F(c^o, p) = \sum_{i\varepsilon T} (V_i - P_i p_i) = V_T - \sum_{i\varepsilon T} P_i p_i$$

$$\geq V_T - \sum_{i\varepsilon T} P_i$$

for all p. Hence (c^o, p^o, F^o) is a solution of the game.

Example 1. (Fixed penalty rate): Let $q_i = q$ for all i. Recall that the returns are numbered so that

$$P_1 \geq P_2 \geq \cdots \geq P_n.$$

Case A: $0 < q \leq 1$, i.e. $0 < V_i/P_i \leq 1$ for all i.

In this case, $|T| = 0$ and we are in Case 1 or Case 2 of Model 1 according as

$$m < q|\overline{T}| \quad \text{or} \quad m \geq q|\overline{T}| = qn.$$

Case A1: $m < q|\overline{T}|$, i.e. $m/n < q$. Let k be the largest integer no greater than m/q, and set $\alpha = m - kq$. Then $k < n$ and by the preceding analysis of Case 1 of Model 1, a solution of the game is given by

$$p_i^{\,o} = q \qquad\qquad i \leq k$$

$$p_{k+1}^{\,o} = \alpha$$

$$p_i^{\,o} = 0 \qquad\qquad i > k+1$$

$$c_i^{\,o} = 1 \qquad\qquad i > k$$

$$c_i^{\,o} = P_{k+1}/P_i \qquad\qquad i \leq k$$

$$F^o(c^o, p^o) = \sum_{i=1}^{n} P_i (q - p_i^{\,o}) c_i^{\,o}$$

$$= P_{k+1}(q - \alpha) + q \sum_{i > k+1} P_i$$

$$= q \sum_{i > k} P_i - \alpha P_{k+1}.$$

We can eliminate α by the substitution $\alpha = m - kq$, yielding

$$F(c^o, p^o) = q\left(\sum_{i > k} P_i + k P_{k+1} \right) - m P_{k+1}.$$

59

Case A2: $m \geq q|\overline{T}|$, i.e. $m/n \geq q$. Then by the preceding analysis of Case 2 of Model 1, a solution of the game is given by

$$p_i^{\ o} = m/n \geq q \qquad \text{for all i}$$

$$c_i^{\ o} = 0 \qquad \text{for all i}$$

$$F(c^o, p^o) = 0 \ .$$

Case B: $q > 1$, i.e. $V_i/P_i > 1$ for all i.

In this case $|\overline{T}| = 0$ and $|T| = n$. Since $m < n$ we are in Case 1, with $k = m$ and $\alpha = 0$. Thus a solution of the game is

$$p_i^{\ o} = 1 \qquad \qquad i \leq m$$

$$p_i^{\ o} = 0 \qquad \qquad i > m$$

$$c_i^{\ o} = 1 \qquad \qquad \text{for all i.}$$

$$F(c^o, p^o) \quad = \sum_{i=1}^{n} (V_i - P_i p_i^{\ o}) c_i^{\ o} = \sum_{i=1}^{n} P_i (q - p_i^{\ o})$$

$$= \sum_{i \leq m} P_i (q-1) + \sum_{i > m} q P_i = V_{tot} - \sum_{i \leq m} P_i \ ,$$

where $V_{tot} = \sum_{i=1}^{n} V_i$.

Example 2 (Two-tier penalty system). Here we assume that $N = \{1,2,\ldots,n\}$ is partitioned into two subsets B and S, of respective sizes $|B|$ and $|S|$, corresponding respectively to "big" and "small" levels of V_i/P_i:

$$q_i = V_i/P_i = q_b \qquad\qquad i\epsilon B$$

$$q_i = V_i/P_i = q_s \qquad\qquad i\epsilon S$$

and $q_b > q_s$.

Case A: $q_b > q_s > 1$. In this case $T = N$ and so $|T| = n$. Since $m<n$, we are in Case 1 of Model 1 with $k=m$ and $\alpha=0$. A solution is given by

$$p_i^o = 1 \qquad\qquad i\leq m$$

$$p_i^o = 0 \qquad\qquad i>m$$

$$c_i^o = 1 \qquad\qquad \text{for all } i$$

$$F(c^o,p^o) = \sum_{i=1}^{n} P_i(q_i-p_i^o)c_i^o$$

$$= \sum_{i=1}^{n} P_i(q_i-p_i^o) = \sum_{i=1}^{n} V_i - \sum_{i=1}^{m} P_i.$$

$$= V_{tot} - \sum_{i=1}^{m} P_i.$$

Case B: $q_b > 1 > q_s$. In this case $T=B$ and $\overline{T}=S$. There are two cases according as $m < |B| + q_s|S|$ or $m\geq|B|+q_s|S|$.

Case B1: $m < |B| + q_s|S|$. This falls under Case 1 of Model 1. With k and α as defined earlier, a solution is given by

$$p_i^o = 1 \qquad\qquad i\epsilon B, \ i\leq k$$

$$p_i^o = q_s \qquad\qquad i\epsilon S, \ i\leq k$$

$$p_{k+1}^o = \alpha$$

$$p_i^o = 0 \qquad\qquad i>k+1,$$

'ogether with

$$c_i^o = 1 \qquad\qquad i\epsilon B, \ i\leq k,$$

61

$$c_i^o = 1 \qquad\qquad i > k$$

$$c_1^o = P_{k+1}/P_i \qquad\qquad i \varepsilon S, \ i \leq k.$$

The resultant payoff is

$$F(c^o, p^o) \ = \ \sum_{\substack{i \varepsilon B \\ i \leq k}} V_i \ + \ \sum_{i > k} V_i \ - \ \sum_{\substack{i \varepsilon B \\ i \leq k}} P_i \ - \ P_{k+1}\alpha.$$

<u>Case B2:</u> $m \geq |B| + q_s|S|$. This is Case 2 of Model 1. A solution is given by strategies satisfying

$$p_i^o = 1 \qquad\qquad i \varepsilon B$$

$$p_i^o \geq q_s \qquad\qquad i \varepsilon S$$

$$c_i^o = 1 \qquad\qquad i \varepsilon B$$

$$c_i^o = 0 \qquad\qquad i \varepsilon S,$$

with payoff

$$F(c^o, p^o) = (q_b - 1) \sum_{i \varepsilon B} P_i.$$

<u>Case C:</u> $1 > q_b > q_s$. Then $\overline{T} = N$. There are two cases according as $m < q_b|B| + q_s|S|$ or $m \geq q_b|B| + q_s|S|$.

<u>Case C1:</u> $m < q_b|B| + q_s|S|$. Then we are in Case 1 of Model 1. A solution is given by

$$p_i^o = q_i \qquad\qquad i \leq k$$

$$p_{k+1}^o = \alpha$$

$$p_i^o = 0 \qquad\qquad i > k+1$$

$$c_i^o = 1 \qquad\qquad i > k$$

$$c_i^o = P_{k+1}/P_i \qquad\qquad i \leq k$$

$$F(c^o, p^o) = \sum_{i > k} V_i - P_{k+1}\ \alpha.$$

<u>Case C2:</u> $m \geq q_b|B| + q_s|S|$. Then we are in Case 2 of Model 1 and a solution is given by the conditions

$$p_i^o \geq q_i \qquad\qquad \text{for all } i$$

$$c_i^o = 0 \qquad\qquad \text{for all } i$$

$$F(c^o, p^o) = 0.$$

Model 2

For this model we use a nonzero-sum game with the two objective functions:

$$F(c,p) = \sum_{i=1}^{n} (V_i - P_i p_i) c_i = \sum_{i=1}^{n} P_i (q_i - p_i) c_i$$

for the taxpayer, and

$$G(c,p) = - \sum_{i=1}^{n} V_i c_i (1 - p_i)$$

for the auditor. (Thus the latter wants to minimize the extent of <u>unde-tected</u> cheating.) It will be convenient to number the returns so that

$$V_1 \geq V_2 \geq \cdots \geq V_n.$$

The solution concept used in this nonzero-sum case is that a pair (c^o, p^o) of strategies is an <u>equilibrium point</u> if c^o maximizes $F(c, p^o)$ over all choices c of the taxpayer's strategy, and p^o maximizes $G(c^o, p)$ over all choices p of the inspector's strategy. (Thus neither player has an incentive to deviate unilaterally from "his" member of the pair (c^o, p^o).)

<u>Case 1.</u> This is defined by $m < |T| + \Sigma\{q_i | i \epsilon \overline{T}\}$. As in Model 1, let k be the largest integer such that

$$\sum_{i=1}^{k} \min\{1, q_i\} \leq m.$$

Then $0 \leq k < n$. Set

$$\alpha = m - \sum_{i=1}^{k} \min\{1, q_i\};$$

it follows that $0 < \alpha \leq \min\{1, q_{k+1}\}$;

<u>Theorem:</u> The pair of strategies (c^o, p^o), where

$$c_i^o = 1 \qquad\qquad i \epsilon T,$$

$$c_i^o = V_{k+1}/V_i \qquad\qquad i \leq k+1, \; i \epsilon \overline{T},$$

$$c_i^o = 1 \qquad\qquad i > k+1, \; i \epsilon \overline{T},$$

$$p_i^o = \min\{1, q_i\} \qquad\qquad i \leq k,$$

$$p_{k+1}^o = \alpha$$

$$p_i^o = 0 \qquad\qquad i > k+1,$$

is an equilibrium point.

Proof: Clearly $\sum_{i=1}^{n} p_i \leq m$ as required. For any strategy c for the taxpayer we have

$$F(c,p^o) = \sum_{i=1}^{n} P_i (q_i - p_i^o) c_i$$

$$= \sum_{\substack{i \epsilon T \\ i \leq k}} P_i (q_i - 1) c_i + \sum_{i > k+1} P_i q_i c_i + P_{k+1}(q_{k+1} - \alpha) c_{k+1}.$$

Since each $P_i(q_i - 1) \geq 0$ for $i \epsilon T$ and since $q_{k+1} - \alpha \geq 0$ and $c_{k+1}^o = 1$, it follows that $F(c^o, p^o) \geq F(c, p^o)$ for all c.
On the other hand,

$$G(c^o, p) = - \sum_{i=1}^{n} V_i c_i^o (1 - p_i)$$

$$= - \sum_{i \epsilon T} V_i (1 - p_i) - \sum_{\substack{i \leq k+1 \\ i \epsilon \overline{T}}} V_{k+1}(1 - p_i) - \sum_{\substack{i > k+1 \\ i \epsilon \overline{T}}} V_i (1 - p_i)$$

$$= - \sum_{i \epsilon T} V_i - V_{k+1} \sum_{\substack{i \leq k+1 \\ i \epsilon \overline{T}}} 1 - \sum_{\substack{i > k+1 \\ i \epsilon \overline{T}}} V_i + \sum_{i \epsilon T} V_i p_i + V_{k+1} \sum_{\substack{i \leq k+1 \\ i \epsilon \overline{T}}} p_i + \sum_{\substack{i > k+1 \\ i \epsilon \overline{T}}} V_i p_i$$

Note that the first three sums do not depend on p. Since the V_i's are in decreasing order of magnitude, it follows that in order to maximize $G(c^o, p)$ we should set

(*) $\qquad\qquad p_i = 1 \qquad\qquad\qquad$ for $i \epsilon T$, $i \leq k$.

This is possible, since by the definition of k, we have $\sum_{\substack{i \epsilon T \\ i \leq k}} 1 \leq m$. Thus, with (*), we have

$$G(c^o, p) = - \sum_{\substack{i \epsilon T \\ i > k}} V_i - V_{k+1} \sum_{\substack{i \leq k+1 \\ i \epsilon \overline{T}}} 1 - \sum_{\substack{i > k+1 \\ i \epsilon \overline{T}}} V_i + \sum_{\substack{i \epsilon T \\ i > k+1}} V_i p_i$$

$$+ V_{k+1} \sum_{\substack{i \leq k+1 \\ i \epsilon \overline{T}}} p_i + \sum_{\substack{i > k+1 \\ i \epsilon \overline{T}}} V_i p_i$$

$$- \sum_{i \geq k+1} V_i - V_{k+1} \sum_{\substack{i < k+1 \\ i \epsilon \overline{T}}} 1 + \sum_{i > k+1} V_i p_i + V_{k+1} p_{k+1}$$

$$+ V_{k+1} \sum_{\substack{i \leq k \\ i \epsilon T}} p_i .$$

64

Note that the coefficient of V_{k+1}, apart from a constant, is $(p_{k+1} + \sum_{\substack{i \leq k \\ i \varepsilon T}} p_i)$. In order to maximize $G(c^o, p)$, this coefficient should be made as large as possible consistent with (*) and the condition $\sum_{i=1}^{n} p_i \leq m$. Writing these restrictions as

$$p_{k+1} + \sum_{\substack{i \leq k \\ i \varepsilon T}} p_i + \sum_{i > k+1} p_i \leq m - \sum_{\substack{i \varepsilon T \\ i \leq k}} 1$$

$$= m - \sum_{\substack{i \varepsilon T \\ i \leq k}} \min\{1, q_i\}$$

$$= \sum_{\substack{i \leq k \\ i \varepsilon T}} q_i + \alpha,$$

we see that p^o satisfies this relation as an equality and makes the coefficient as large as possible. Thus p^o maximizes $G(c^o, p)$ and (c^o, p^o) is an equilibrium point.

The resulting values of F and G are

$$F^o = \sum_{\substack{i \varepsilon T \\ i \leq k}} V_i + \sum_{i \geq k+1} V_i - \sum_{\substack{i \varepsilon T \\ i \leq k}} P_i - \alpha P_{k+1} ,$$

$$G^o = -\sum_{i \geq k+1} V_i + V_{k+1}(m-k).$$

Case 2. This is defined by $m \geq |T| + \Sigma\{q_i | i \varepsilon \overline{T}\}$.

Theorem: Any pair of strategies (c^o, p^o) where

$$c_i^{\ o} = 1 \qquad\qquad i \varepsilon T,$$

$$c_i^{\ o} = 0 \qquad\qquad i \varepsilon \overline{T},$$

$$p_i^{\ o} \geq \min\{1, q_i\} \qquad\quad \text{for all } i,$$

is an equilibrium point.

Proof: For any c, we have

$$F(c, p^o) = \sum_{i=1}^{n} P_i(q_i - p_i^{\ o})c_i$$

$$= \sum_{i \varepsilon T} P_i(q_i - 1)c_i + \sum_{i \varepsilon \overline{T}} P_i(q_i - p_i^{\ o})c_i^{\ o}.$$

For $i \varepsilon T$, since $P_i(q_i-1) > 0$, it follows that in order to maximize $F(c,p^o)$ we must have $c_i = 1$. Also, since $P_i(q_i - p_i^o) \leq 0$ for $i \varepsilon \overline{T}$, it follows that

$$F(c,p^o) \leq F(c^o,p^o) \qquad \text{for all } c.$$

On the other hand,

$$G(c^o,p) = - \sum_{i=1}^{n} V_i c_i^o (1-p_i)$$

$$= - \sum_{i \varepsilon T} V_i (1-p_i).$$

Hence $G(c^o,p) \leq G(c^o,p^o) = 0$ for all p and we have shown that (c^o,p^o) is an equilibrium point. The resultant values of F and G are

$$F^o = \sum_{i \varepsilon T} V_i - \sum_{i \varepsilon T} P_i,$$

$$G^o = 0.$$

APPENDIX C

This attachment presents graphical representations of the relationship between discriminant function scores and tax change for each of the audit classes. The correlation coefficients derived from this data are discussed on page 10 of the text.
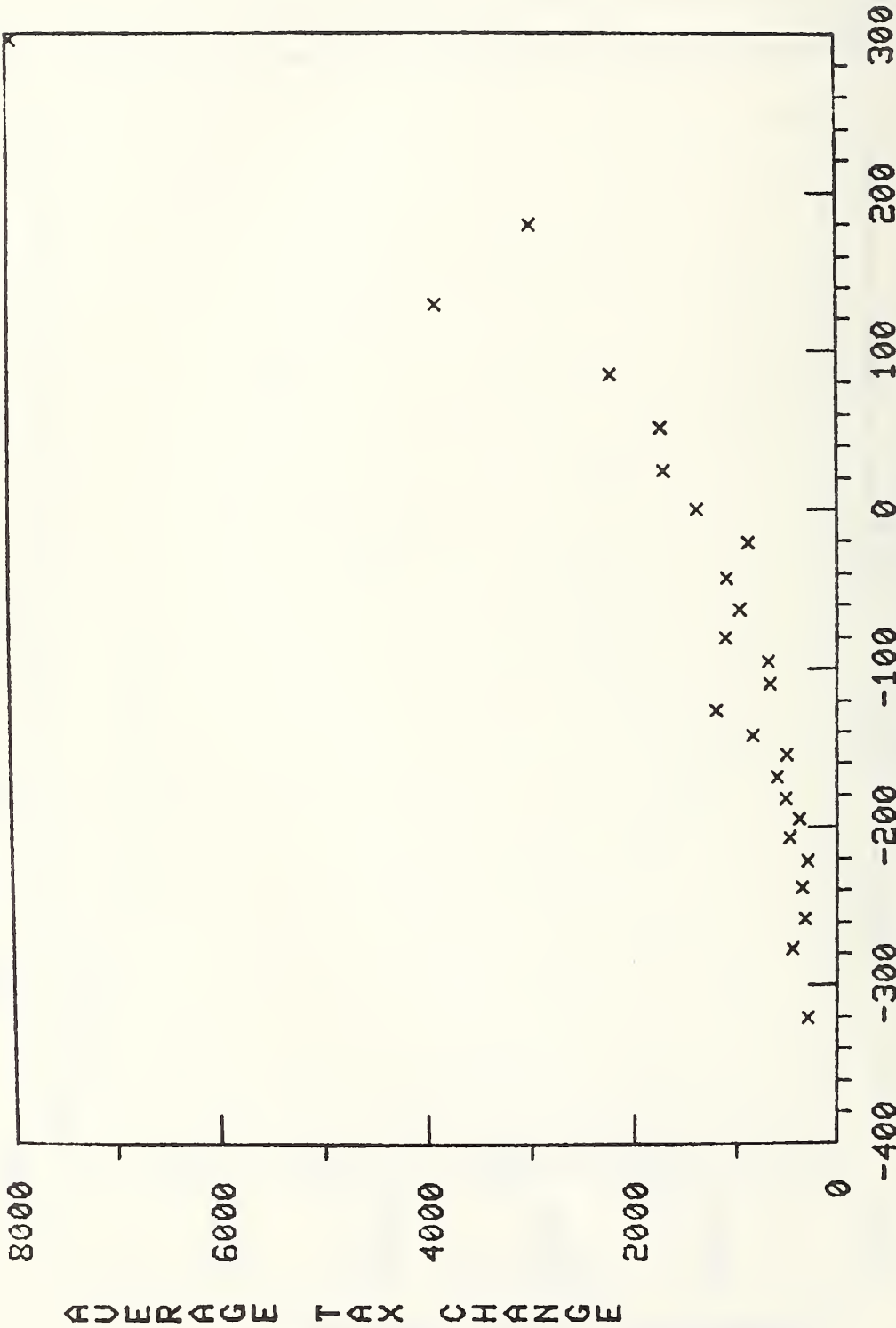
AVG DIF SCORE VS AVG TAX CHANGE FOR LOW BUSINESS CLASS
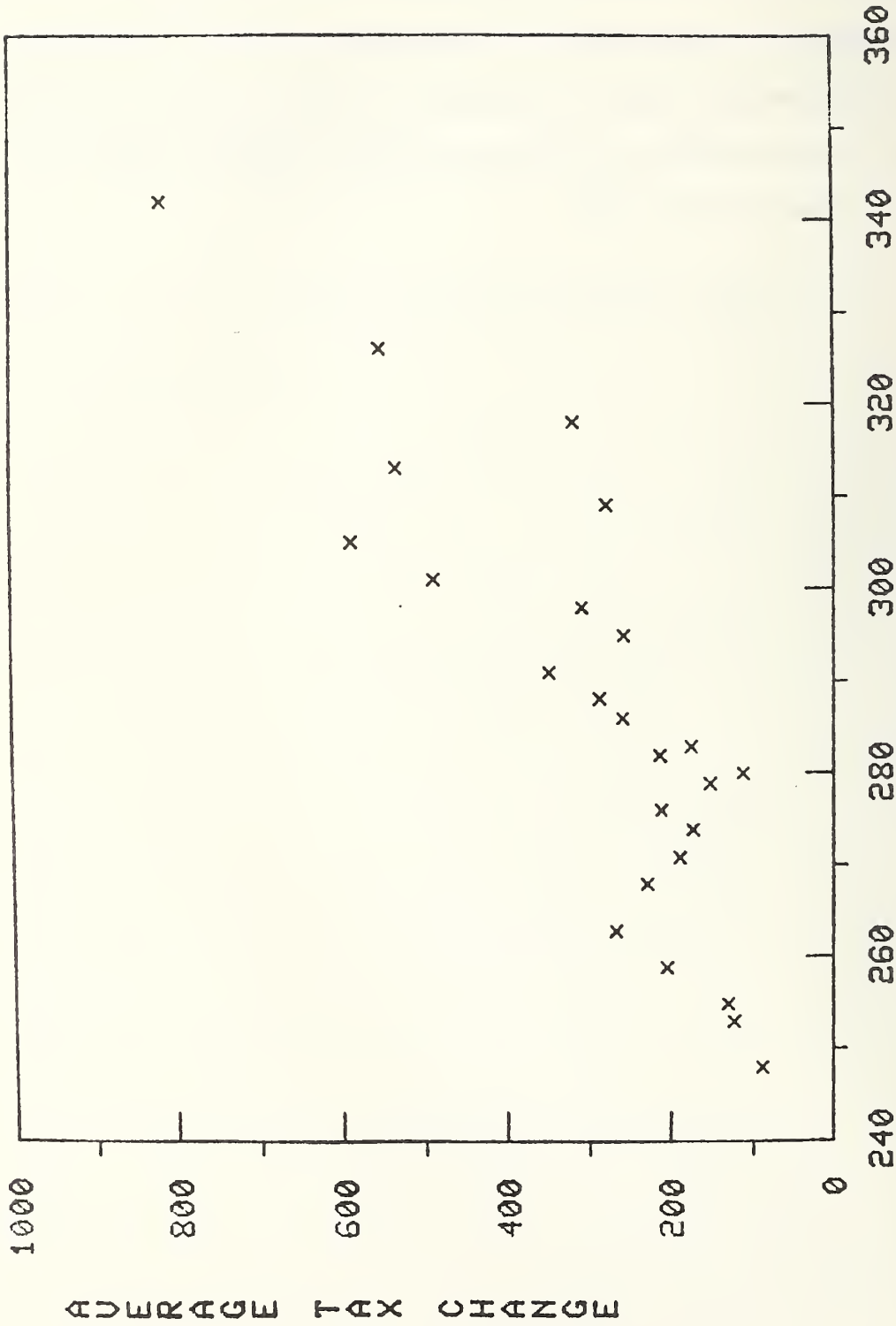
TCMP SAMPLE OF RETURNS FILED IN 1966
AVERAGE DIF SCORE
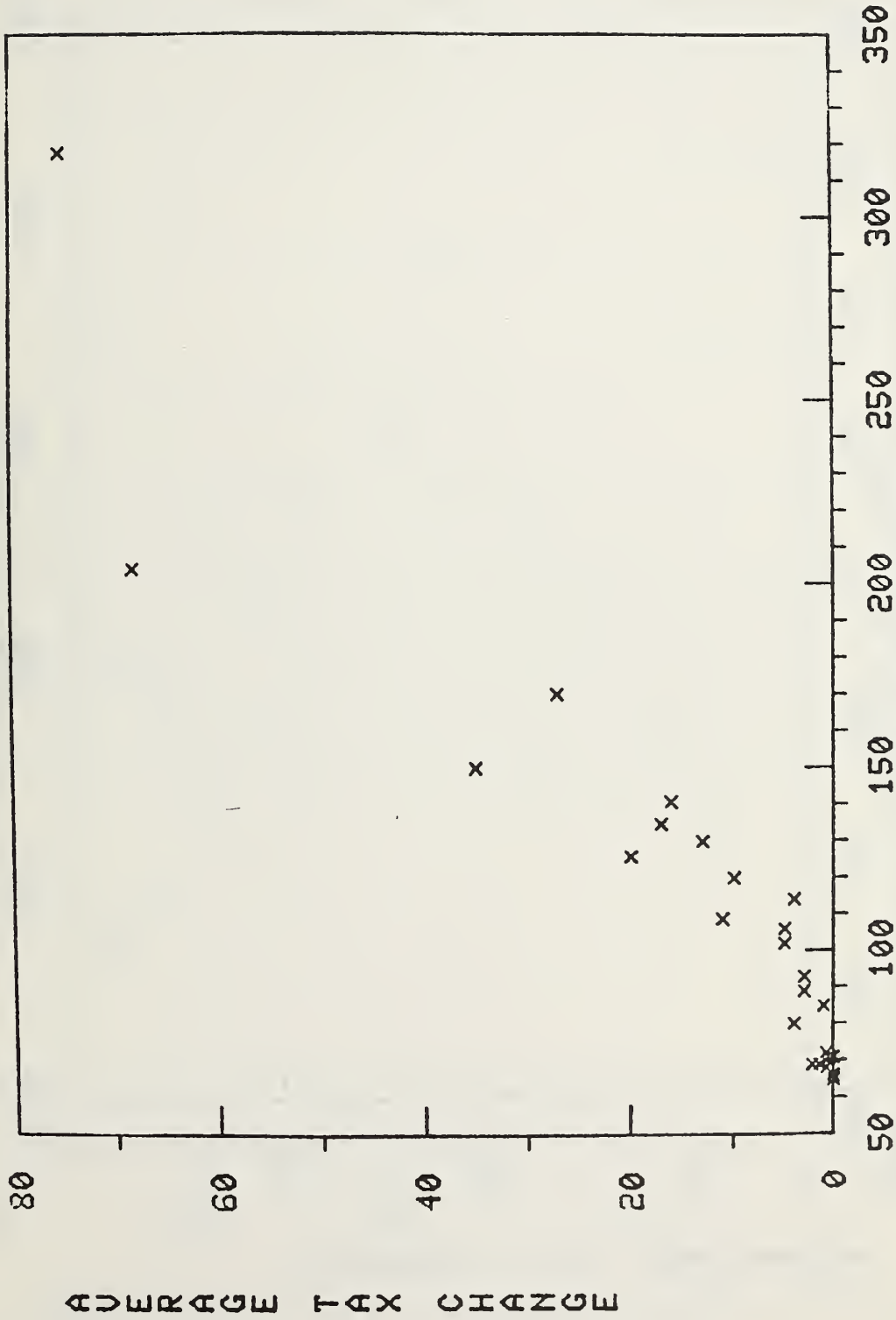EACH POINT REPRESENTS 4% OF POPULATION (RANKED BY DIF SCORE)

FIGURE C1

AVG DIF SCORE VS AVG TAX CHANGE FOR MEDIUM BUSINESS CLASS
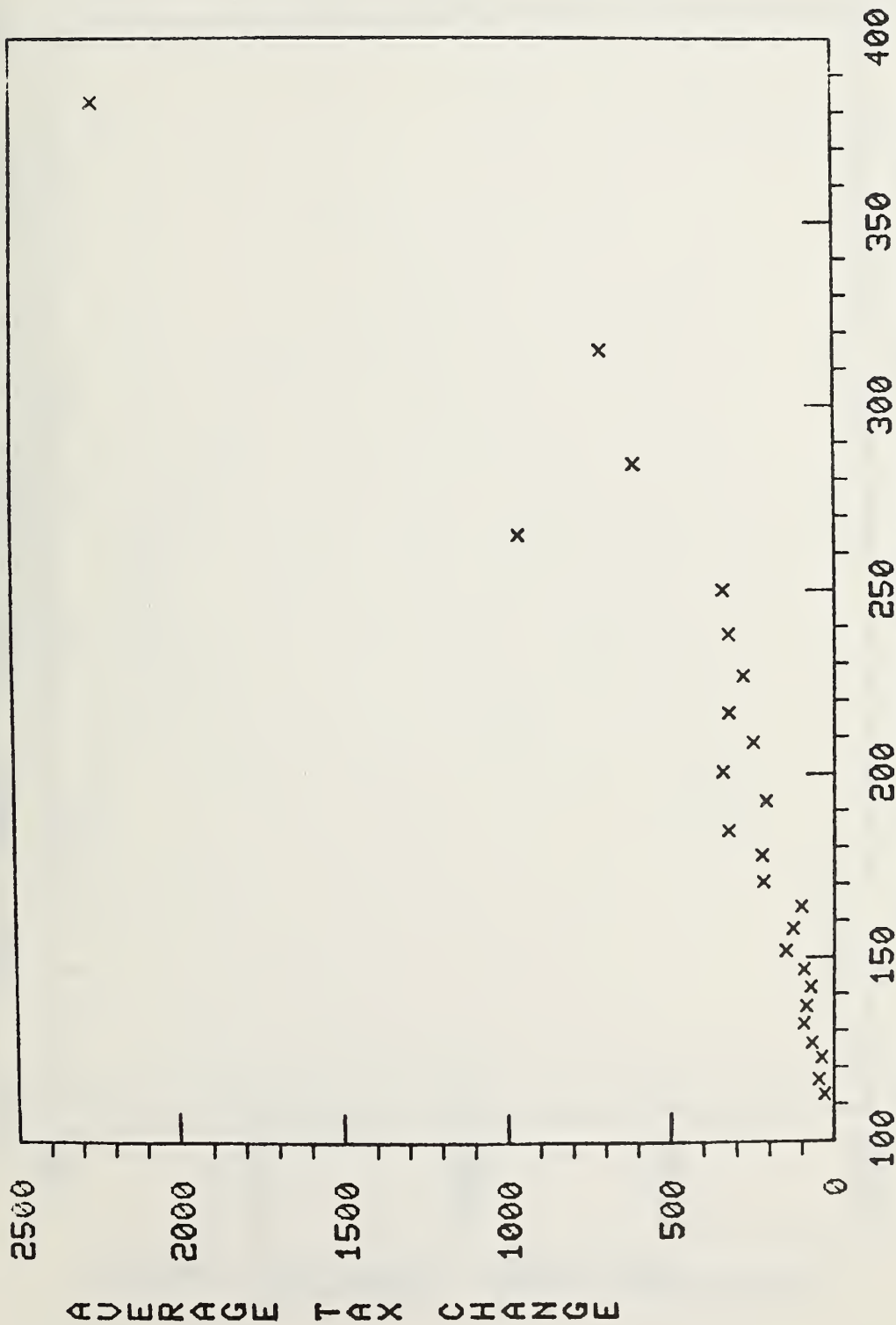
TCMP SAMPLE OF RETURNS FILED IN 1966
AVERAGE DIF SCORE
EACH POINT REPRESENTS 4% OF POPULATION (RANKED BY DIF SCORE)

FIGURE C2

AVG DIF SCORE VS AVG TAX CHANGE FOR HIGH BUSINESS CLASS

AVERAGE DIF SCORE
TCMP SAMPLE OF RETURNS FILED IN 1966
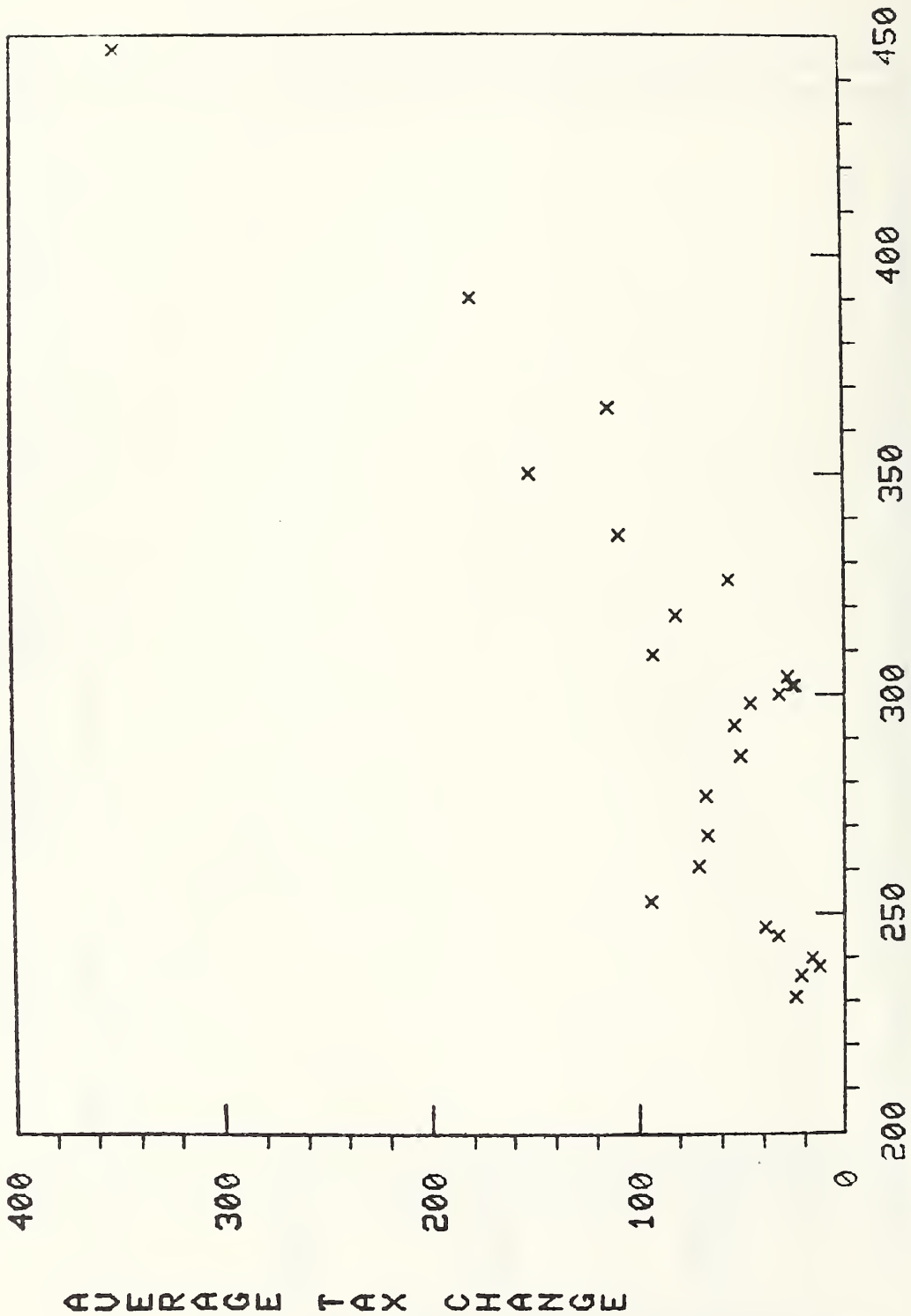EACH POINT REPRESENTS 4% OF POPULATION (RANKED BY DIF SCORE)

FIGURE C3

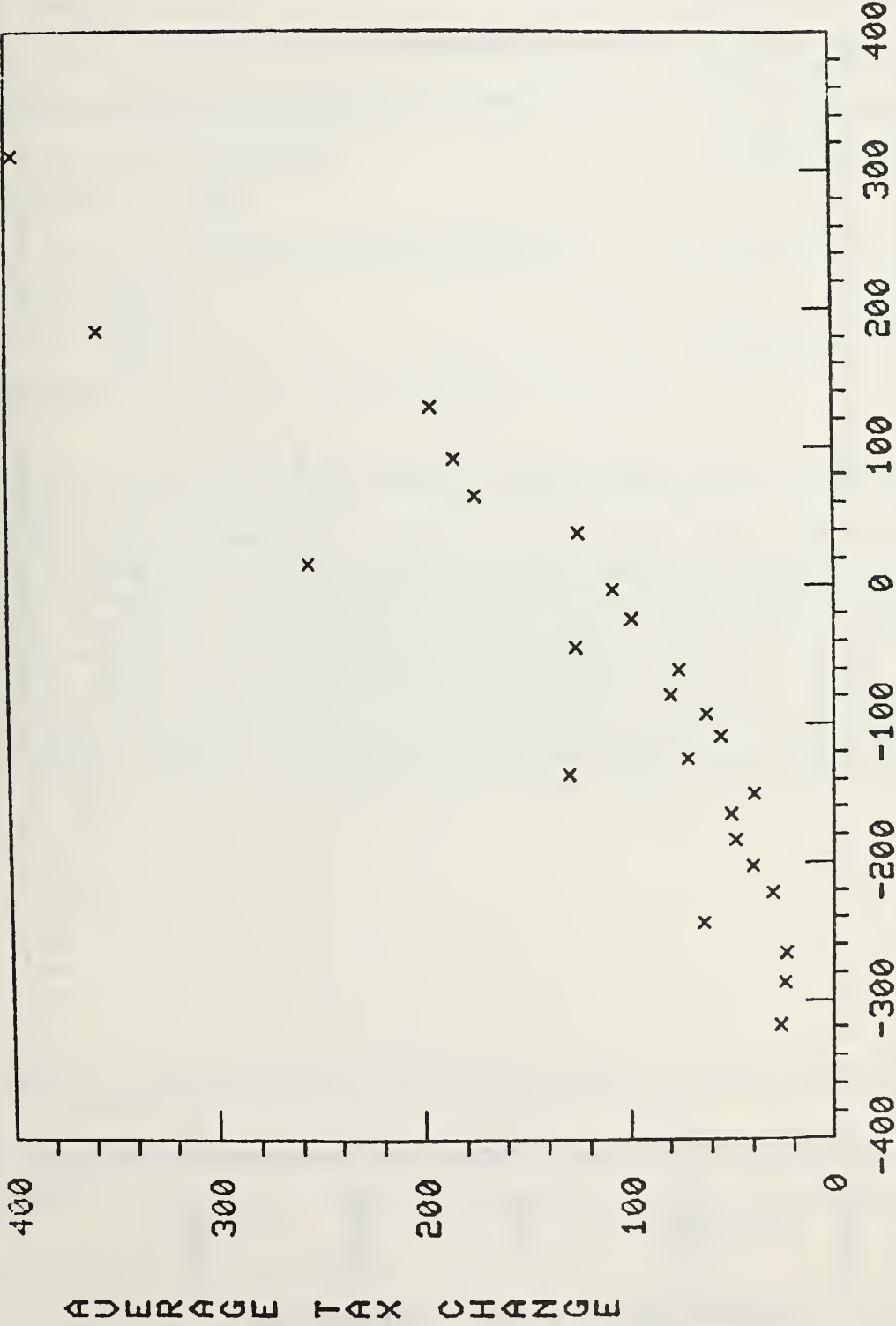AVG DIF SCORE VS AVG TAX CHANGE FOR 1040A RETURNS

TCMP SAMPLE OF RETURNS FILED IN 1966

AVERAGE DIF SCORE

EACH POINT REPRESENTS 4% OF POPULATION (RANKED BY DIF SCORE)

FIGURE C4

71

AVG DIF SCORE VS AVG TAX CHANGE FOR $10,-$15,000 AGI RETURNS

TCMP SAMPLE OF RETURNS FILED IN 1966
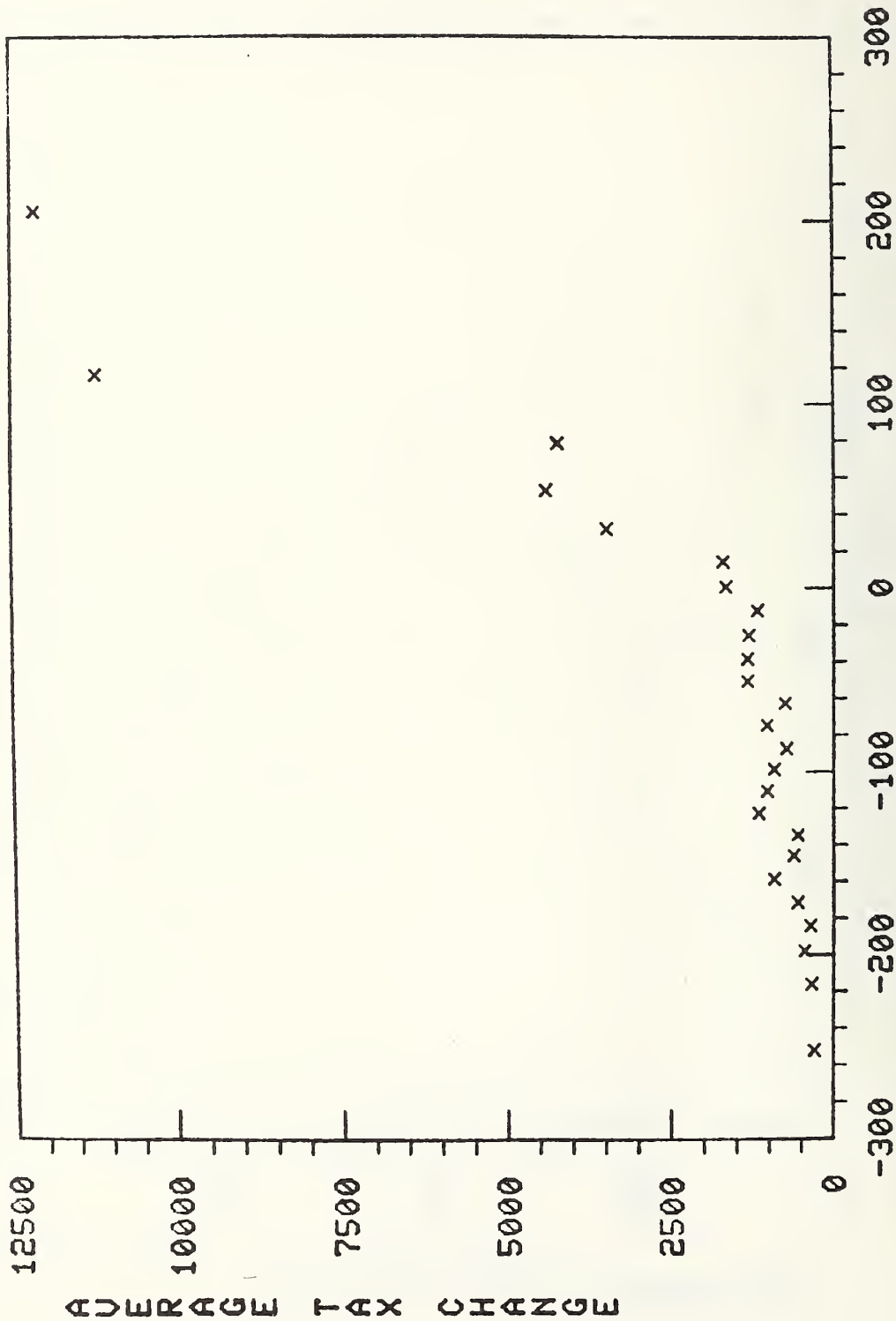AVERAGE DIF SCORE
EACH POINT REPRESENTS 4% OF POPULATION (RANKED BY DIF SCORE)

FIGURE C5

72

AVG DIF SCORE VS AVG TAX CHANGE FOR $15,-50,000 AGI RETURNS

TCMP SAMPLE OF RETURNS FILED IN 1966
AVERAGE DIF SCORE
EACH POINT REPRESENTS 4% OF POPULATION (RANKED BY DIF SCORE)

FIGURE C6

73

AVG DIF SCORE VS AVG TAX CHANGE FOR ITEM. RETURNS >$50,000

AVERAGE DIF SCORE

TCMP SAMPLE OF RETURNS FILED IN 1966

EACH POINT REPRESENTS 4% OF POPULATION (RANKED BY DIF SCORE)

FIGURE C7

74

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO.	2. Gov't Accession No.	3. Recipient's Accession No.

4. TITLE AND SUBTITLE

Resource Requirements and Allocations in IRS' Audit Division ; A Systems Analysis

5. Publication Date

6. Performing Organization Code

7. AUTHOR(S)
Karla L. Hoffman, Lambert S. Joel, Martin H. Pearl

8. Performing Organ. Report No.

9. PERFORMING ORGANIZATION NAME AND ADDRESS

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

10. Project/Task/Work Unit No.
7120413

11. Contract/Grant No.

12. Sponsoring Organization Name and Complete Address *(Street, City, State, ZIP)*
Internal Revenue Service
1111 Constitution Avenue
Washington, D.C. 20224

13. Type of Report & Period Covered

14. Sponsoring Agency Code
IRS

15. SUPPLEMENTARY NOTES

16. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)*

The Applied Mathematics Division of the National Bureau of Standards (NBS) was asked by the Internal Revenue Service under IRS order #7T2867 to "assess the validity and effectiveness of the IRS Audit Division's Long Range Plan's strategies and approaches to resource requirements and allocations". This report, which documents that assessment, (1) summarizes the examination and evaluation by NBS of IRS's current audit practices and plans for the future, and (2) presents our major conclusions and recommendations. In many cases the available information did not permit recommending specific methods with which to solve current problems. In such situations we sought to identify those areas in which we believe further research is needed and is most likely to lead to improvements over present practices.

17. KEY WORDS *(six to twelve entries; alphabetical order; capitalize only the first letter of the first key word unless a proper name; separated by semicolons)*

Audit; behavior; compliance; discriminant analysis; game theory; law enforcement; taxation; utility

18. AVAILABILITY	☐ Unlimited	19. SECURITY CLASS (THIS REPORT)	21. NO. OF PAGES
☐ For Official Distribution. Do Not Release to NTIS		UNCLASSIFIED	
☐ Order From Sup. of Doc., U.S. Government Printing Office Washington, D.C. 20402, SD Cat. No. C13		20. SECURITY CLASS (THIS PAGE)	22. Price
☐ Order From National Technical Information Service (NTIS) Springfield, Virginia 22151		UNCLASSIFIED	