JATA
Vol. 18, No. 1
Spring 1996
pp 72–91

A Taxpayer Compliance Application of Benford's Law

Mark J. Nigrini

ABSTRACT

This study investigates whether the nonrandom element of human behavior could facilitate the detection of tax evasion. Unplanned Evasion (UPE) is defined to be blatant manipulation by the taxpayer of line items at filing time. Planned Evasion (PE) is the result of planned actions to conceal an audit trail. UPE requires that the taxpayer invent a number(s) for the line item(s).

Benford's Law (Benford 1938) is used as an expected distribution for the digits in tabulated data. The assumption is that the digits of data that are truthfully reported, or are subject to PE, should conform to the expected digital frequencies. A Distortion Factor model that quantifies the extent of UPE is developed. Tax returns on the U.S. Internal Revenue Service Individual Tax Model Files are analyzed. The analysis, based on digital frequencies, indicates that Low Income taxpayers practice UPE to a greater extent than High Income taxpayers.

The U.S. Internal Revenue Service (IRS) estimated that the individual gross tax gap (the difference between income tax due and income tax voluntarily paid) for 1992 would amount to \$91–94 billion (IRS 1990). Jackson and Milliron (1986, 15) describe the state of compliance understanding as "primitive" and Roth et al. (1989, 247) believe that our base of compliance knowledge is "small compared with the importance of the phenomenon." More recently Alm and McKee (1992, 107) claim that despite a decade of empirical analysis "our understanding of compliance remains surprisingly limited." Long and Swingen (1991, 664) stress a need to focus on specific types of noncompliance. The objective of this study is to introduce and identify a type of evasion described as Unplanned Evasion (or filing-time evasion), to increase our understanding of the evasion phenomenon.

Building on the work of Slemrod (1985), recent studies by Christian and Gupta (1993) and Dusenbury (1993) take a new approach to compliance research by drawing inferences from unaudited taxpayer data. Similarly, the basis of this study is that the

Mark J. Nigrini is Assistant Professor of Accounting at Saint Mary's University.

I wish to acknowledge the useful comments of Bruce Busta, Thomas E. Buttross, Charles W. Christian, C. Bryan Cloyd, Julie H. Collins, J. Sanley Fuhrmann, Ted P. Hill, Martin S. Levy, R. Lorraine McClenny, Linda J. Mittermaier, Ralph A. Raimi, Dan L. Schisler, George Schmelzle, and Wallace R. Wood on earlier versions of this paper. I also wish to thank Sandra S. Kramer and Edmund Outslay in their capacities as editors, and the reviewers for their contributions. An earlier version of this paper was presented at the annual meeting of the American Accounting Association in New York City in 1994.

Submitted: May 1992
Accepted: October 1995

digital frequencies of numbers reported in tax returns will deviate from the *expected* frequencies due to taxpayer evasion. The *expected* frequencies are those that would be observed in the event of truthful reporting, and are assumed to be governed by Benford's Law (Benford 1938). This idea is not new in this study, however, it has been applied to investigate whether corporate managers are involved in manipulating reported earnings (Carslaw 1988; Thomas 1989). This study demonstrates that (1) a Distortion Factor model can be used to estimate the extent of evasion for a particular field, and (2) the application of Benford's Law could aid the IRS in estimating the extent of evasion shortly after all returns for a period have been filed. The research questions that are addressed are:

1. Is there a relationship between the digits of the numbers used on tax returns, and tax evasion?
2. If so, can the relationship provide an increased understanding of evasive behavior by indicating the income or deduction fields most prone to evasion?

The remainder of this paper is organized into six sections. In the first section Benford's Law is reviewed. In the second section, a Distortion Factor model that quantifies the level of manipulation in a data set is developed. The data preparation procedure is described in the third section. The fourth section explains the results of applying the model to taxpayer data. The fifth and sixth sections contain a discussion and conclusions.

BENFORD'S LAW

Benford's Law (Benford 1938) is a somewhat counter-intuitive property of the digits appearing in tabulated numerical data. The naive expectation is that each digit is equally likely to be the first digit in tabulated data; hence, an expected probability of 1/9 for any of 1 through 9 as the first (or leading) digit. With ten possible second digits, the naive expectation is that each of 0 through 9 has an equal probability of 1/10 of being the second digit. Benford's Law shows that the digits of tabulated data are heavily skewed towards the lower digits for digits in the first position. A (smaller) bias in favor of the lower digits also exists for the second position. Under Benford's Law the expected digital frequencies follow a logarithmic pattern. The formulas are shown below with D_1 representing the first, and D_2 the second digit of a number. A two-digit combination is written as D_1D_2. The formulas for the expected digital frequencies are (base 10 logarithms):

$$\mathrm{P}(D_1 = d_1) = \log(1 + (1/d_1)); \qquad d_1 \in \{1, 2, ..., 9\} \tag{1}$$

$$\mathrm{P}(D_2 = d_2) = \sum_{d_1=1}^{9} \log(1 + (1/d_1d_2)); \qquad d_2 \in \{0, 1, ..., 9\} \tag{2}$$

$$\mathrm{P}(D_1D_2 = d_1d_2) = \log(1 + (1/d_1d_2)) \tag{3}$$

$$\mathrm{P}(D_2 = d_2|D_1 = d_1) = \log(1 + (1/d_1d_2))/\log(1 + (1/d_1)) \tag{4}$$

where P indicates the probability of observing the event in parentheses. The probabilities can also be written as a General Significant-Digit Law (Hill 1995) where, for example,

$$\mathrm{P}(D_1D_2D_3 = 147) = \log(1 + (1/147)) \cong 0.0029$$

The expected frequencies of Benford's Law are shown in table 1.

Benford's Law applies to lists of numbers that describe the relative sizes of similar phenomena. Examples include the market values of listed corporations, populations of counties in the U.S., and the sizes of files on a computer's hard drive. The list should

TABLE 1
BENFORD'S LAW: EXPECTED DIGITAL FREQUENCIES

	Position in Number			
Digit	**1st**	**2nd**	**3rd**	**4th**
0		.11968	.10178	.10018
1	.30103	.11389	.10138	.10014
2	.17609	.10882	.10097	.10010
3	.12494	.10433	.10057	.10006
4	.09691	.10031	.10018	.10002
5	.07918	.09668	.09979	.09998
6	.06695	.09337	.09940	.09994
7	.05799	.09035	.09902	.09990
8	.05115	.08757	.09864	.09986
9	.04576	.08500	.09827	.09982

Note: The number 147 has three digits, with a 1 as the first digit, 4 as the second digit, and a 7 as the third digit. The table indicates that under Benford's Law the expected proportion of numbers with a first digit 1 is 0.30103 and the expected proportion of numbers with a third digit 7 is 0.09902.

not have an arbitrary maximum or minimum value. Benford's Law would not apply to assigned numbers such as Social Security numbers, car license plate numbers or telephone numbers, for example.

Prior Accounting Applications

The theme behind most applications of Benford's Law is that it provides a negative test of naturalness. That is, conformity of a data set to Benford's Law does not necessarily imply naturalness, but nonconformity should raise some level of suspicion (Nigrini 1994). Carslaw (1988) focused on the second digits of reported income and relied on a cognitive emphasis on the first digit to detect evidence of income manipulation. Numbers just below an $a10^k$ (a,k integer) boundary were postulated to be valued abnormally less than a number above this boundary. For example, an amount of $798,000 (or $19.96 million) would be seen to be abnormally less than $803,000 (or $20.03 million). These $a10^k$ numbers might form the targets to be attained by management, thereby providing an incentive to report income above a reference point. Evidence of $a10^k$ reference points would be an excess of 0s and a shortage of 9s as the second digit in reported income numbers. Carslaw compared the digital frequencies of income numbers of New Zealand companies to Benford's Law.

Thomas (1989) analyzed U.S. COMPUSTAT data, Earnings Per Share (EPS) numbers, and quarterly data. He tabulated the first two-digits of Earnings Before Extraordinary Items and these results also showed evidence of income manipulation. With respect to the rightmost digit of EPS numbers, there were excesses of 0s and 5s, and a shortage of 9s for positive earnings.

Christian and Gupta (1993) analyzed taxpayer data to examine secondary evasion. Secondary evasion is inferred from their data and occurs when taxpayers reduce taxable income from above a tax table income bracket to below a tax table bracket (Slemrod

1985). Christian and Gupta (1993) use Benford's Law to justify the assumption that the ending two-digits of taxable income should be uniform over the [00,99] range. Their analysis showed that a higher than expected percentage of tax table users were positioned in the upper dollars of the table brackets.[1]

In July 1995, the digital frequencies of selected tax return data were analyzed by the Dutch Ministry of Finance. The objective of the analysis was to detect noncompliance, to assess the feasibility of incorporating selected digital tests in an audit selection model, and to assess whether third party reporting contributes to compliance. A sample of 30,000 tax returns for fiscal 1992 were analyzed. The analysis included a tabulation, with graphs, of the interest received data per taxpayer, according to the third party returns from Dutch banks to the taxation authorities. The first and second digit frequencies showed a near-perfect conformity to Benford's Law.

THE DISTORTION FACTOR MODEL

An excess of lower first digits suggests a downward manipulation of true numbers, but the extent of the manipulation is unknown. To estimate the level of distortion (or manipulation) in an income or deduction field requires a comparison of the mean of the actual numbers and the mean of the numbers in a Benford Set. Data sets are called Benford Sets if they closely approximate Benford's Law. However, there is no unique mean for the numbers in a Benford Set because a Benford Set could comprise relatively small, or relatively large, numbers. The lack of a unique mean problem was solved by moving the decimal point of each actual number (where necessary) so that the number falls into the range (10,100) (also written as $10 \leq X < 100$). For example, an actual number of 110,364 is *collapsed* to 11.0364, and an actual number of 2.204 is *expanded* to 22.04. With actual numbers scaled to the (10,100) range, a comparison can be made to the mean of a set of numbers conforming to Benford's Law scaled to the same (10,100) range.

The basis of Benford's Law is that when the data are ordered (ranked from smallest to largest) the data form a geometric sequence. Benford (1938, 563), after citing a number of examples of geometric patterns in natural phenomena, declared that "Nature counts e^0, e^x, e^{2x}, e^{3x},... and builds and functions accordingly." Lemons (1986) proves that a geometric sequence will follow Benford's Law. Raimi (1976) notes that even sequences (such as the Fibonacci numbers) that are approximately geometric, and almost all sequences defined by linear recursions, will follow Benford's Law.[2]

A geometric sequence can be written as ar^{n-1}, where a is the first element of the sequence, and r is the ratio of the $(n + 1)$th element divided by the nth element. The ratio, r, of a sequence is a function of the range (e.g., (10,100)) and N (the number of observations) and is computed as follows:

$$r = 10^{(\log(ub) - \log(lb))/N} \tag{5}$$

[1] There is, however, no way to distinguish evaders from taxpayers who became more diligent to find legitimate deductions when near a table boundary. Dusenbury (1993) analyzes secondary evasion from a prospect theory perspective.

[2] A geometric sequence is a set of numbers in ascending order where each number is a constant multiple of the preceding number (e.g., each number is 1.01 times the preceding number). A sequence can be uniquely defined by a starting value, an ending value, the common ratio, and a specified number of observations.

where ub and lb represent the upper bound and lower bound of the geometric sequence respectively. Although the range spans the half-open interval (10,100), the upper bound is taken to be 100 for the purposes of computing r. Using Equation (5) to compute r, the largest element of the sequence will tend to the upper bound (in this case 100) as N tends to infinity.

For the sequence of *collapsed* values, a is fixed at 10, and "$\log(ub) - \log(lb)$" is fixed at 1. The mean of the sequence is dependent on N (r being a function of N) but is only slightly affected by N for N greater than about 500. The equation for the expected mean (*EM*) of the observations of a Benford Set (scaled to the (10,100) range) is computed in appendix A. The *EM* of a large Benford Set with all numbers scaled to the (10,100) range is approximately 39.08. If all digital combinations had an equal chance of occurring the mean would be 55 [(100 + 10)/2]. The *EM* is less than that expected for uniformly distributed (10,100) numbers because of the higher probability associated with the lower digits. The distortion in a data set of reported income tax numbers (whole dollars only) is computed using the following steps:

1. Transform reported numbers to numbers in the range (10,100):
   a. Delete all numbers that are less than 10 (this includes all numbers reported as zero dollars). This step ensures that all numbers have an explicit first and second digit.
   b. *Collapse* all reported numbers equal to or above 100 to the range (10,100) by moving the decimal point as required.[3]
2. Compute the Actual Mean (*AM*) of the *collapsed* numbers.
3. Compute the *EM* of the observations of a Benford Set scaled to the (10,100) range using Equation (A2).
4. Compute the Distortion Factor (*DF*):

$$DF = (AM - EM)/EM \tag{6}$$

The *DF* (multiplied by 100) measures the percentage deviation of the *AM* from the *EM*. An excess of lower first digits indicates that more smaller numbers were used compared to a Benford Set and the *DF* would be negative. The second and later digits influence the *DF*, but the effect is less because a manipulation of these digits is a smaller percentage manipulation.

Limitations of the DF Model

Pinkham (1961) addressed the question of how digital frequencies would be affected if the data were all multiplied by the same number (a non-zero constant). He proved that the Benford frequencies are the only set of frequencies that are invariant under a change of scale. Under Pinkham's scale invariance theorem, consistent percentage manipulations, upwards or downwards, throughout a Benford Set will not be detectable by testing digital frequencies. Scale invariance causes the *actual* frequencies (after manipulation) to be the same as the *expected* frequencies. Minor changes in the digital frequencies may occur in small data sets. Benford's Law can therefore not be used to detect manipulations that are systematic in this way. Raimi (1969, 118) shows that if a set of numbers is not a Benford Set, then multiplication by any k might move the set further away or closer to the *expected* frequencies. However, no k exists such that multiplication would cause a non-conforming data set to become a Benford Set.

[3] The choice of the range does not affect the Distortion Factor if step (2) and step (3) use the same range.

With respect to the digital frequencies of tax data, the assumption is that the digital frequencies of evasion-free data should approximate those of a Benford Set. Also, the meaning of evasion needs to be refined to identify specifically the behavior apparent from the digital frequencies of income or deduction fields. Cuccia (1994, 108) discusses redefining the reporting decision. He notes that most compliance research has considered the reporting decision as being made *at the time of filing*, unrelated to any other decisions or underlying transaction. He states that reporting decisions faced by taxpayers are much more complicated, involving numerous decisions at various points in time and subject to different consequences. After reviewing some limitations inherent in compliance research, he notes that care should be taken by researchers to explicitly define the nature of the underreporting being examined.

Due to the scale invariance property, it is necessary to consider which types of evasive behaviors will be detected by the *DF* model. The broad concept of evasion is therefore divided into two types of evasion dependent mainly on timing and skill. Compliance usually means that the taxpayer accurately reports the tax liability and that the return is filed at the proper time. Selected behavioral tax literature (Cowell 1985; Carroll 1989; Kidder and McEwen 1989; Klepper and Nagin 1989) suggests that the following dichotomy of tax evasion is plausible:

1. *Planned Evasion (PE)*: The intention to evade exists early in the fiscal year and the taxpayer takes preplanned steps to hide an audit trail and/or obtain fraudulent supporting documents.
2. *Unplanned Evasion (UPE)*: Evasion occurs while the return and supporting statements are being prepared. These evasive actions are a blatant adjustment (downwards for income items and upwards for deduction items) of items thought to be safe and unlikely to be detected prior to an audit, although detectable upon an audit.

The essential difference between PE and UPE is timing, and differences in opportunity, skill, and motivation might cause a taxpayer to be more amenable to a particular form of evasive behavior. The difference between the taxpayer's true tax liability and reported tax liability can therefore be ascribed to PE, UPE, and errors.

With respect to interest received, it is submitted that each taxpayer could be placed into one of the following groups:

1. Taxpayers correctly (honestly) reporting all interest;
2. Taxpayers committing UPE; and
3. Taxpayers using PE methods such as not reporting interest received from a foreign bank.

The scale invariance theorem refers to multiplying *all* the elements by the *same* non-zero constant. It follows that if the elements of a randomly drawn subset (which should also form a Benford Set) were all multiplied by a constant, the complete set of data would again form a Benford Set. A randomly drawn subset of a Benford Set would form a Benford Set both *before* and *after* multiplication. After reinserting the manipulated Benford Set in the original data one would again have a Benford Set. Many randomly drawn subsets could therefore also be drawn from the original data with each being multiplied by a non-zero constant. After reinserting the manipulated subsets back into the original data (that portion not being put into any subset) one would again have a Benford Set.

The third PE group (above) could be divided into many subgroups based on the percentage differences between the true numbers and the reported numbers. If the true interest numbers of each subgroup form a Benford Set, then under the scale invariance theorem, so do the reported numbers because each subgroup's true numbers have been manipulated downwards by approximately the same percentage. Because multiplication by a constant does not affect any particular subgroup (due to the scale invariance property), the reported numbers of the taxpayers in group (3) will form a Benford Set.

UPE is a behavioral act where the taxpayer consciously fabricates a number on the tax return. This act is influenced by the way in which we think of numbers. Carslaw (1988) showed that $a10^k$ (a,k integer) numbers form psychological boundaries in accounting situations and that the 10^k (k, integer) boundary was a "key reference point." Certain numbers are also reference points for consumers, as witnessed by 99 as the ending digits of the prices of many consumer products. Rosch (1975, 532) defines a reference point as a stimulus that other stimuli are seen "in relation to." His experiments showed that subjects used numbers such as 10, 100, and 1,000 as reference points. The UPE assumption is that a reasonable level of evasion is perceived to be a manipulation such that the reported number is in the same $(10^k,10^{k+1})$ (k, integer) range as the true number.[4] The range assumption is motivated by the belief that taxpayers have little knowledge of the "average" return filed by persons in similar economic positions, and that the true number forms an anchor upon which to base the reported number.

The range assumption means that taxpayers just above a 10^k dollar amount will avoid manipulating the income line item and use another line item, or not practice UPE. All manipulated *income* numbers will have *lower* digital combinations than the true values. Income numbers just above 10^k boundaries are most limited in the percentage downward manipulation, and deduction numbers just below 10^k boundaries are most limited in the percentage upward manipulation.

The *DF* model would fail if the UPE reference point assumption were not true, *and* evasion-free numbers followed Benford's Law. Statistically insignificant *DF*s would occur due to chance alone and the sign of the *DF* would be randomly positive or negative. Significant *DF*s, signed consistently with the manipulation incentive for a field, would signal that UPE is occurring *and* that the reference point assumption is valid.

DATA PREPARATION

The 1985 and 1988 Individual Tax Model Files (ITMFs) were tested for signs of UPE.[5] Each ITMF contains about 150 "dollar amount" fields per taxpayer. The DF model was tested on a set of non-business fields amenable to UPE and a set of business (Schedule C) fields *not* amenable to UPE. Taxpayers were classified as having a high or low incentive to commit UPE, and the two groups were analyzed separately. Fields were eliminated that were an unlikely target of UPE using the following criteria:

1. Fields subject to third party reporting (a few exceptions were made);
2. Fields that were the sum or difference of other line items (again, a few exceptions were made);

---

[4] The characteristic of the log (base 10) of the true number is unchanged, only the mantissa is manipulated upwards or downwards.

[5] The ITMFs are compiled by the IRS from a stratified sample of unaudited individual income tax returns (Forms 1040, 1040a, and 1040EZ) filed by U.S. citizens and residents. The 1985 sample of 108,840 returns was selected from the population of 101.7 million returns filed in 1986. The 1988 sample of 95,713 returns was selected from the 109.7 million returns filed in 1989. The populations include all returns processed except for tentative and amended returns. Returns were assigned to sample strata based on the size of total income or loss amounts, business receipts, and the presence or absence of various forms. Returns are selected from the strata at rates ranging from 0.02% to 100%. The sampling method, methods of disguising data to preserve privacy, and a listing of the indicator and amount fields is given in IRS (1985, 1988).

3. Fields that were "blurred" due to IRS identity protection techniques; and
4. Fields subject to statutory maximums (e.g., child care credit and Keogh contributions).

Taxpayers were partitioned into "higher" or "lower" incentive for UPE based on criteria believed to affect audit probability. Fischer et al. (1992) note that unless audits are perfect in their ability to detect noncompliance, the probability of an audit and the probability of detection are not identical. From a behavioral perspective the appropriate operationalization is the *perceived* probability of detection. Recall, however, that UPE is a blatant adjustment at filing time and consequently should be detected in the event of an audit. The general partitioning criteria used were the level of Adjusted Gross Income (AGI), the presence of losses, and the presence of other abnormalities (e.g., filed late or outside of U.S.). The lower incentive group generally had either an AGI above $100,000 or a zero marginal tax rate. Admittedly, some taxpayers may have moved into the zero marginal tax rate subset by using UPE methods. IRS (1991) confirms a higher audit probability for returns with relatively higher Total Positive Income.

The lower incentive group is called the High Income (HI) Group and the higher incentive group is called the Low Income (LO) Group. The AGI amount classified as "high" corresponds to the classification in Christian and Gupta (1993, 80) and is less than the high income cutoff in IRS (1988, 2). The amount is also conveniently a 10^k (k, integer) number to avoid the cutoff amount potentially affecting the DF model. For example, a lower AGI amount might have caused the dollar amounts in the fields analyzed to be too clustered around a low mean to conform to Benford's Law. The partitioning criteria are set out in appendix B.[6] To test the data for UPE, the *DF* model was tested for classification accuracy (i.e., does the *DF* indicate whether the field was overstated or understated?).

RESULTS

The first test for evasion is a simple tabulation of digital frequencies (taking Benford's Law as a maintained hypothesis) and an analysis of the results. The digital frequencies of Interest Received and Total Interest Paid Deduction on the 1985 and 1988 ITMFs were analyzed. These fields were chosen because the data items were similar since each related to interest either being received or paid. The evasion distortion would be that interest paid numbers are overstated and that interest received numbers are understated.

The digital frequencies are shown in tables 2 and 3. The analysis uses the unweighted data which ignore the frequency of sampling from the various strata. Observations below $10 were deleted as these numbers have only one (possibly rounded) digit. The bias column contains a "+" sign if the *actual* frequency exceeds the *expected* (i.e., Benford)

[6] After the field deletion criteria and the HI and LO partitioning criteria, the 1988 ITMF was divided into 63,454 HI and 32,259 LO observations, and 30 fields were deemed amenable to UPE (manipulatable). The 1985 ITMF was partitioned into 71,740 HI and 37,100 LO observations, and 25 fields were deemed manipulatable.

Descriptive statistics of the manipulatable fields in each group were computed as a preliminary check for biases in the data that could affect the digital frequencies and perhaps lead to erroneous conclusions about the existence and extent of UPE. For each field, the mean, standard deviation, maximum value reported, and sample skewness measure were tabulated. The manipulatable fields were those remaining after some fields were deleted due to statutory maximum or minimum dollar restrictions. The descriptive statistics provided no reason to believe that Benford's Law would not be applicable to the remaining data.

TABLE 2
DIGITAL FREQUENCIES OF INTEREST RECEIVED

Panel A: First Digit Frequencies

First Digit	Expect	1985 ITMF Actual	1985 ITMF Bias	1985 ITMF Z-Stat	1988 ITMF Actual	1988 ITMF Bias	1988 ITMF Z-Stat
1	.3010	.3041	+	2.039*	.3059	+	2.958*
2	.1761	.1786	+	1.943	.1779	+	1.326
3	.1249	.1252	+	0.224	.1270	+	1.740
4	.0969	.0970	+	0.051	.0948	−	1.971*
5	.0792	.0779	−	1.396	.0778	−	1.444
6	.0669	.0665	−	0.532	.0650	−	2.143*
7	.0580	.0568	−	1.490	.0563	−	1.998*
8	.0512	.0493	−	2.550*	.0503	−	1.119
9	.0458	.0446	−	1.641	.0450	−	1.005

Panel B: Second Digit Frequencies

Second Digit	Expect	1985 ITMF Actual	1985 ITMF Bias	1985 ITMF Z-Stat	1988 ITMF Actual	1988 ITMF Bias	1988 ITMF Z-Stat
0	.1197	.1214	+	1.553	.1239	+	3.625*
1	.1139	.1133	−	0.595	.1145	+	0.496
2	.1088	.1104	+	1.505	.1107	+	1.692
3	.1043	.1032	−	1.125	.1031	−	1.131
4	.1003	.1004	+	0.069	.0996	−	0.626
5	.0967	.0969	+	0.255	.0964	−	0.304
6	.0934	.0925	−	0.930	.0936	+	0.251
7	.0904	.0895	−	0.885	.0891	−	1.241
8	.0876	.0871	−	0.473	.0857	−	1.855
9	.0850	.0854	+	0.438	.0835	−	1.530

*Significant at the 0.05 level.
Note: Observations total 91,022 for 1985 and 78,640 for 1988. The expected proportions are those of Benford's Law. The Bias column reads + if the actual proportions exceed those of Benford's Law, and − otherwise. The Z-Statistic is calculated per Equation 7.

frequency, and "−" sign if the converse is true. The Z-statistic tests the null hypothesis that the *actual* proportion equals the *expected* proportion of Benford's Law.[7]

The interest received digital frequencies in table 2 show that for the lower first digits, the actual frequencies exceed the expected frequencies. Conversely, the actual frequencies of the higher first digits are less than the expected frequencies. The Z-statistics indicate that only two of the first digit differences in 1985, and four of the first digit

[7] The Z-statistic is calculated as follows:

$$Z = (|p_o - p_e| - 1/2N)/\sqrt{(p_e \times (1 - p_e)/N)} \tag{7}$$

where p_e denotes the expected proportion, p_o the observed proportion, and N the number of observations. The $(1/2N)$ term is a continuity correction term and is only used when it is smaller than the first term in the numerator.

TABLE 3
DIGITAL FREQUENCIES OF TOTAL INTEREST PAID

Panel A: First Digit Frequencies

First Digit	Expect	1985 ITMF Actual	1985 ITMF Bias	1985 ITMF Z-Stat	1988 ITMF Actual	1988 ITMF Bias	1988 ITMF Z-Stat
1	.3010	.2963	−	2.737*	.2901	−	5.563*
2	.1761	.1729	−	2.197*	.1654	−	6.533*
3	.1249	.1280	+	2.436*	.1236	−	0.921
4	.0969	.0973	+	0.325	.1025	+	4.420*
5	.0792	.0821	+	2.916*	.0841	+	4.248*
6	.0669	.0674	+	0.447	.0711	+	3.917*
7	.0580	.0577	−	0.289	.0627	+	4.740*
8	.0512	.0523	+	1.429	.0548	+	3.872*
9	.0458	.0459	+	0.185	.0455	−	0.248

Panel B: Second Digit Frequencies

Second Digit	Expect	1985 ITMF Actual	1985 ITMF Bias	1985 ITMF Z-Stat	1988 ITMF Actual	1988 ITMF Bias	1988 ITMF Z-Stat
0	.1197	.1206	+	0.755	.1238	+	2.971*
1	.1139	.1142	+	0.256	.1114	−	1.823
2	.1088	.1086	−	0.204	.1087	−	0.083
3	.1043	.1009	−	2.979*	.1039	−	0.310
4	.1003	.1005	+	0.165	.1013	+	0.739
5	.0967	.0988	+	1.896	.0953	−	1.060
6	.0934	.0941	+	0.641	.0927	−	0.537
7	.0904	.0892	−	1.045	.0907	+	0.252
8	.0876	.0874	−	0.185	.0853	−	1.857
9	.0850	.0858	+	0.715	.0869	+	1.579

*Significant at the 0.05 level.
Note: Observations total 70,725 for 1985 and 54,737 for 1988. The expected proportions are those of Benford's Law. The Bias column reads + if the actual proportions exceed those of Benford's Law, and − otherwise. The *Z*-Statistic is calculated per Equation 7.

differences in 1988 are significant at the 0.05 level.[8] For second digits the excesses are mainly associated with the lower digits and the shortfalls are mainly associated with the higher digits.

The interest paid deduction digital frequencies in table 3 show that for the lower first digits, the actual frequencies are less than the expected frequencies. The opposite is evident for the higher digits. Exceptions are the 1985 first digit 7 and the 1988 first digit 9 (these differences have insignificant *Z*-statistics). For the second digits, no trend is apparent from the bias signs.

The excess of the lower first digits for interest income suggests that some lower numbers were reported than should have been reported. In contrast, the excess of the

[8] The absolute differences are small and the significance is largely due to the power from the large sample size.

high first digits for interest paid suggests that more high numbers were reported than should have been reported. The direction of the deviations is consistent with tax evasion behavior to the extent that Benford's Law holds in its application to taxation data.

To test the *DF* model, the *EM* of the 1985 interest received data (table 2) was computed as follows:

$$EM = 90/(91022 \times (10^{(1/91022)} - 1))$$
$$= 39.0860.$$

The *AM* of the collapsed values was 38.6756 and consequently,

$$DF = (38.6756 - 39.0860)/39.0860$$
$$= -0.0115.$$

The *DF* shows that the reported numbers are 1.15 percent below those that would be found in a Benford Set. For the 1988 interest received data, the *AM* was 38.5972 and the *EM* was 39.0859, which resulted in a *DF* of −0.0125.

For the 1985 and 1988 interest paid fields, the *AM*s were 39.3318 and 40.0202, respectively. In both cases the *AM*s exceeded the *EM*s, and the *DF*s were computed to be .0063 and .0239 for 1985 and 1988 respectively. The digital combinations (which suggested that the interest received fields were understated and the interest paid fields were overstated), and the signs of the *DF*s (which were negative for the interest received fields and positive for the interest paid fields) are therefore consistent. To conduct a statistical test to investigate whether a *DF* is indeed significant, requires the standard deviation of the *DF* [*SD*(*DF*)] which is computed in appendix A, equation (A10).

Table 4 presents the set of 1985 *DF*s. The expected sign of a field (given some UPE) is negative for income and gain fields, and positive for deduction and loss fields.[9] The table has two sections, the first listing income and adjustment items on the first page of the Form 1040 and the second listing itemized deductions. The proportion of fields for which the *DF* is correctly signed for the LO Group is 14/25 (56.0%). The incorrectly signed *DF*s are mainly for deduction fields and fields used by a low number of taxpayers. The proportion of LO Group *DF*s correctly signed and significant ($p < 0.05$) is 10/17 (58.8%). The proportion of fields correctly signed is lower for the HI Group at 11/25 (44.0%).

Table 5 presents the 1988 field *DF*s. The results are similar to those for 1985, in that the proportion of fields for which the *DF* is correctly signed for the LO Group is 19/30 (63.3%). Again the incorrectly signed *DF*s are mainly for deduction fields and fields used by a few taxpayers. The proportion of LO Group *DF*s correctly signed and significant ($p < 0.05$) is 9/13 (69.2%). The proportion of fields correctly signed is lower for the HI Group at 13/30 (43.3%).

Table 6 summarizes the classification accuracy of the field *DF*s. In panel A it can be seen that for each year the proportion correctly signed (i.e., signed in a manner consistent with UPE) is higher for the LO Group than for the HI Group. Also, for those fields where the *DF* is significantly different from zero, the proportion correctly signed is highest for the LO Group. The classification accuracy is more pronounced for the 1988 sample.

---

[9] The expected sign for cash contributions under $3,000 is negative because numbers above 1,000 (and below 3,000) would have low first digits that would bias the *DF* negatively.

TABLE 4
DISTORTION FACTORS PER FIELD: 1985 INDIVIDUAL TAX MODEL FILE

	LO Group		HI Group		
Field Description	*DF*	N	*DF*	N	**Exp. Sign**
Interest Received	−0.029*	26310	−0.003	64712	−
Dividends Received	−0.037*	8519	−0.009*	47502	−
Refund state/local tax	−0.004	9993	−0.012*	27567	−
Business income (Sch C)	0.008	5019	−0.021*	13092	−
Capital gains	−0.019	3685	0.013*	38434	−
Other gains	−0.102*	227	0.008	6450	−
Pensions/annuities	0.013	4696	−0.012	8496	−
Taxable pension/annuity	0.066*	704	0.001	1589	−
Supplemental income	0.024*	3334	0.010*	20369	−
Farm profits	−0.058	211	0.006	827	−
Unemployment comp.	−0.094*	2106	−0.012*	814	−
Moving expense	0.048*	862	0.041*	1337	+
Employee bus. expense	−0.001	3965	0.019*	9649	+
Forfeited int. penalty	−0.108*	280	−0.000	799	+
Total medical/dental	−0.121*	7930	−0.044*	16531	+
General sales tax	0.104*	19869	0.136*	55269	+
Personal property tax	−0.073*	6005	−0.021*	21809	+
Other taxes	−0.121*	4353	−0.196*	11275	+
Mortgage: fin. instit.	0.033*	15178	−0.014*	39370	+
Total charitable cont.	0.006	18971	−0.014*	54668	+
Cash under $3,000	−0.018*	18572	−0.019*	53323	−
Cash over $3,000	0.155*	846	0.016*	13812	+
Other than cash	−0.107*	5988	−0.048*	22367	+
Net casualty/theft	−0.076	103	−0.090	176	+
Total itemized deduct.	0.134*	20455	−0.002	56453	+

*Significant at the 0.05 level
Note: The Distortion Factor is $(AM - EM)/EM$, where *AM* is the Actual Mean and *EM* the Expected Mean of the collapsed observations. Total Low Income taxpayers = 37,100 and High Income = 71,740. The Expected Sign is negative for income/gain items and positive for deduction/loss items. The Expected Sign for "Cash under $3,000" is negative because observations between $1,000 and $3,000 have low collapsed values.

Panels B and C of table 6 summarize the field *DF* results in contingency tables. Panel B indicates that for 1985 and 1988 the proportion of fields correctly signed is 60 percent for the LO Group and 43.6 percent for the HI Group. The Chi-square test for independence has a test statistic of 2.949 (1 degree of freedom) which has a significance level of 0.106. Therefore it cannot be concluded that there is a statistical dependence between classification accuracy and taxpayer Group. However, the evidence does not contradict the hypothesis of a relationship between UPE and Group membership since the proportion of fields correctly signed is higher for the LO Groups. Fisher's exact test (one-tail) shows the probability of observing a table that gives at least as much evidence of association as the one actually observed, given that the null hypothesis is true (SAS

TABLE 5
DISTORTION FACTORS PER FIELD: 1988 INDIVIDUAL TAX MODEL FILE

Field Description	LO Group *DF*	LO Group N	HI Group *DF*	HI Group N	Exp. Sign
Interest Received	−0.022*	22452	−0.009*	56188	−
Dividends Received	−0.016*	6627	−0.014*	41109	−
Refund state/local tax	0.009	7401	−0.001	26844	−
Business inc. (Sch C)	−0.014	3915	0.043*	15445	−
Capital gains	0.023	2263	−0.017*	30246	−
Supplemental net gains	−0.027	154	−0.010	5663	−
Taxable IRA distrib.	−0.056*	882	−0.049*	1966	−
Pensions/annuities	−0.013	5395	−0.004	13306	−
Taxable pension/annuity	−0.012	5178	−0.004	10955	−
Farm income	−0.049	387	0.009	1081	−
Unemployment comp.	−0.067*	2103	−0.056*	1267	−
Employee business exp.	0.009	546	0.027	1683	+
Forfeited int. penalty	0.007	207	−0.027	772	+
Self-employ. health ins.	0.090*	561	0.217*	4773	+
Total itemized deduct.	0.202*	12251	−0.004	47822	+
Prescription/medical	0.008	1255	0.016	1694	+
Medical and dental	−0.043	789	−0.071*	915	+
Total interest paid	0.158*	11551	−0.012*	43186	+
Mortgage: fin. instit.	0.141*	10163	−0.008*	35103	+
Deductible points	−0.049*	974	−0.017	4584	+
Investment interest	−0.037	238	−0.003	18700	+
Personal interest	−0.051*	10368	−0.022*	32139	+
Total charitable cont.	0.008	11356	−0.021*	46003	+
Nonlimited misc. ded.	−0.059	276	−0.045*	1335	+
Noncash contributions	−0.120*	4294	−0.045*	16781	+
Cash contributions	−0.011	11163	−0.021*	45639	+
Net casualty/theft	−0.234*	40	−0.000	75	+
Moving expenses	0.030	355	0.030	1055	+
Employee business exp.	−0.018	3150	−0.013	7359	+
Tax preparation fee	0.073*	3324	−0.021*	13861	+

*Significant at the 0.05 level
Note: The Distortion Factor is $(AM - EM)/EM$, where AM is the Actual Mean and EM the Expected Mean of the collapsed observations. Total Low Income taxpayers = 32,259 and High Income = 63,454. The Expected Sign is negative for income/gain items and positive for deduction/loss items.

1985, 413). The Fisher exact-test probability is 0.063, which is somewhat short of a 0.05 significance level. Panel C relates to those fields where the *DF*s are significantly different from zero. While the percentage correctly signed is greater for the LO Group (when compared to panel B) and approximately equal to that of the HI Group (in panel B), the test statistics indicate a lower level of significance due to the smaller sample size.

Panel C supports the ability of the *DF* model to detect UPE based on a LO Group accuracy rate of 63.3%. Since the percentage of field *DF*s correctly signed is higher for the LO Groups, it seems that detection probability affects the incidence of UPE. How-

TABLE 6
SUMMARY OF FIELD DISTORTION FACTOR CLASSIFICATION ACCURACY

Panel A: Proportion of fields correctly signed
1985:

	Number of Fields	Proportion Correct
Low Income Group		
All Fields	25	0.560
Fields with significant *DF*s	17	0.588
High Income Group		
All Fields	25	0.440
Fields with significant *DF*s	17	0.529

1988:

	Number of Fields	Proportion Correct
Low Income Group		
All Fields	30	0.633
Fields with significant *DF*s	13	0.692
High Income Group		
All Fields	30	0.433
Fields with significant *DF*s	16	0.375

Panel B: Distortion Factors of all fields (1985 and 1988)

	Number of Fields	Proportion Correct	Proportion Incorrect
Low Income Group			
All Fields	55	0.600	0.400
High Income Group			
All Fields	55	0.436	0.564

Chi-Square: 2.949 (probability = 0.086)
Fisher's Exact Test (one-tail): probability = 0.063

Panel C: Distortion Factors of all fields significantly different from zero (1985 and 1988)

	Number of Fields	Proportion Correct	Proportion Incorrect
Low Income Group			
All Fields	30	0.633	0.367
High Income Group			
All Fields	33	0.455	0.545

Chi-Square: 2.022 (probability = 0.155)
Fisher's Exact Test (one-tail): probability = 0.121

Note: The table presents the classification accuracy of the Distortion Factors (*DF*s) shown in tables 4 and 5 in a contingency table format.

ever, the percentages are not significantly different due in part to the small sample size. If the cell counts in panels B and C were doubled by, for example, analyzing 1986 and 1987 data and obtaining the same level of classification accuracy, the Chi-square statistics would all be significant at the 0.05 level.

The *DF*s of fields *not* amenable to UPE were tested for classification accuracy. Recall that if the reference point assumption was not valid, the *DF*s would be randomly positive and negative, and probably statistically insignificant. The 1988 ITMF contains 13 Schedule C fields. Roth et al. (1989, 60) report that only 37.8% of taxpayers reporting Schedule C income had the net amount corrected by $50 or less in the 1982 TCMP cycle. These line items are therefore subject to widespread manipulation which should conform to the definition of PE, as taxpayers are unlikely to complete the Schedule based on invented numbers. The more likely behavior is to hide income or pad expenses, and then to report the fraudulent result. Since the reported numbers are not invented, one-half of the *DF*s should be signed correctly due to chance alone.

Table 7 indicates that only three of the 13 Schedule C fields (23.1%) have *DF* signs consistent with those expected under UPE. A test for an equal number (13/2) of correct and incorrect classifications gives a Chi-square test statistic of 3.77 ($p = 0.062$). The results support the position that the *DF* captures the extent and direction of UPE (and not PE). None of the correctly signed field *DF*s are significantly different from zero. The Schedule C fields are therefore manipulated to the extent that the reference point assumption is violated.

DISCUSSION

The *DF* model could be used by the IRS to detect UPE shortly after the returns for a year have been filed. The most recent TCMP was for 1988. A 1995 TCMP is being planned (*The Wall Street Journal* 1995), but evasion estimates are only calculable after a few years. Widely spaced TCMP audits are not effective at capturing trends in tax return fields. In contrast, the *DF* model is relatively inexpensive and can be used to formulate UPE estimates shortly after all returns for a year have been filed. Use of the model would require that the ITMFs be analyzed in conjunction with TCMP data bases.

TABLE 7
DISTORTION FACTORS PER FIELD: 1988 SCHEDULE C FIELDS

Field Description	*DF*	Observations	Expected Sign
Gross Income	0.032*	25360	−
Net Sales	0.025*	24657	−
Cost of Goods Sold	0.003	7418	+
Total Deductions	−0.005	24941	+
Car and Truck Expenses	−0.022*	13434	+
Depreciation	−0.015*	13922	+
Commissions	−0.022	2459	+
Mortgage Interest	−0.012	2925	+
Other Interest	0.010	5903	+
Office Expenses	−0.024*	12125	+
Insurance	−0.014*	11597	+
Rent	−0.024*	7572	+
Wages (net of jobs credit)	0.002	5876	+

*Significant at the 0.05 level
Note: The Distortion Factor is $(AM - EM)/EM$, where *AM* is the Actual Mean and *EM* the Expected Mean of the collapsed observations. The number of taxpayers using each field is shown. Total 1988 ITMF taxpayers = 95,713. The Expected Sign is negative for income/gain items and positive for deduction/loss items.

With a relationship between TCMP detected evasion (PE and UPE) and the field *DF*s (which measure UPE only), evasion estimates could be made shortly after the filing deadline. An issue for future research would be the formulation of a revenue loss estimate due to UPE. This would encompass a composite *DF* for the population of taxpayers and the multiplication by an appropriate tax rate.

The *DF* model assumes that UPE (expressed as a percentage of the true number) is constant across the number ranges ((10,100) etc.), and that taxpayers committing UPE use 10^k (k, integer) numbers as reference points. These assumptions might be confirmed with TCMP data. A confirmation could be problematic as the TCMP file only has details of the reported and corrected amounts. It might not be possible to conclude which evasive acts were planned and which were unplanned. Another approach could be to survey taxpayers willing to discuss their evasive actions.

The methodology could be used on corporate returns. The theoretical challenge would be to reformulate the expected effect of evasion on digital frequencies because the blatant adjustments defined as UPE are not expected to occur on corporate returns. The detection of fraud on corporate returns could assist the IRS in reducing the corporate tax gap and the theory may assist with other applications. The IRS could possibly use digital frequency and *DF* model tests to check the validity of third party reporting by different banks and other entities required to file third party reports.

The *DF* model might also be suitable for the detection of fraud or other irregularities. Auditing researchers could identify situations where an assessment needs to be made of the integrity of tabulated data such as lists of accounts receivable, accounts payable or inventory counts. Researchers could compare actual digital frequencies to Benford's Law and use the *DF* model to assess whether "human intervention" was likely. A limitation of the research is that the literature has relatively few papers in which a number of different data sets are tested for conformity to Benford's Law. Such tests would alert users to possible threats to its validity.

SUMMARY AND CONCLUSION

The objective of the study was to establish a link between tax evasion and the numbers reported by taxpayers. Evasion was described as either Planned Evasion (PE) or Unplanned Evasion (UPE). PE was said to occur when the intention to evade existed early in the fiscal year and the taxpayer took steps to hide an audit trail. UPE was said to occur while the return was being prepared.

Benford's Law provided expected frequencies for the digits in tabulated data. The hypothesis was that the human element inherent in fabricated (fraudulent) numbers would cause the digits to deviate from the expected frequencies. A Distortion Factor (*DF*) Model was developed that allows the direction and extent of manipulation in a data set to be quantified. In general, when the *actual* proportion of lower digits exceeds the *expected* frequencies, the model indicates that the numbers have been manipulated downward.

The *DF* model detected UPE on the 1985 and 1988 Individual Tax Model Files. On the basis that income items would be understated and deduction items would be overstated, the *DF*s indicated that a higher proportion of Low Income taxpayer fields were correctly signed when compared to the *DF*s of High Income taxpayers. The *DF*s of fields not amenable to UPE (but known to contain PE) were computed to check the signaling validity of the *DF* model. Only a few Schedule C (business income) *DF*s were signed in a manner consistent with UPE. The *DF* model could be used by the IRS to detect UPE shortly after the returns for a year have been filed.

Benford's Law holds promise as an interesting and useful area of tax evasion and auditing research. Researchers evaluating data sets will need to assess whether the expected digital frequencies are valid for the data under consideration, how fraud might occur, and how the fraud could affect the digital frequencies.

APPENDIX A
THE MEAN AND VARIANCE OF THE DISTORTION FACTOR

The expected mean and variance of the *DF* is based on a geometric sequence spanning the range (10,100). A geometric sequence can be written as ar^{n-1}, where a is the first element of the sequence, r is the ratio of the $(n + 1)$th element divided by the nth element, and N is the number of observations.

The *DF* measures the deviation of the *AM* (Actual Mean) from the *EM* (Expected Mean) to determine whether the values of the numbers appears to be over or understated relative to a set of numbers that conforms to Benford's Law. Since the actual numbers are scaled to the (10,100) range, the *AM* is compared to the *EM* of a Benford Set of numbers scaled to the same range. The *EM* and variance of a geometric sequence spanning the (10,100) range is derived below.

The Expected Sum (*ES*) of the elements of any geometric sequence is,

$$ES = a \times (r^N - 1)/(r - 1). \tag{A1}$$

The *EM* of elements spanning the (10,100) range is derived by substituting $a = 10$, $r = 10^{1/N}$ (equation 5), and then dividing by N,

$$EM = 90/[N \times (10^{1/N} - 1)]. \tag{A2}$$

The variance (V) is now calculated using the computational formula for the variance of a random variable (X), where,

$$\mathrm{V}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2. \tag{A3}$$

Concentrating on the first term ($\mathrm{E}(X^2)$) in (A3), let the sequence be denoted S_n, then,

$$(S_n)^2 = a^2 r^{(2n-2)}; \tag{A4}$$

$$\begin{aligned} \mathrm{E}(S_n)^2 &= (\Sigma\, a^2 r^{(2n-2)})/N \\ &= 9900/(N \times (r^2 - 1)). \end{aligned} \tag{A5}$$

Returning to (A3) and substituting (A5) and (A2),

$$\begin{aligned} \mathrm{V}(X) &= \frac{9900}{N \times (r^2 - 1)} - \frac{90^2}{(N \times (r - 1))^2} \\ &= \frac{900 \times (N11r - N11 - 9r - 9)}{N^2 \times (r^3 - r^2 - r + 1)}. \end{aligned} \tag{A6}$$

$\mathrm{V}(X)$ is the variance of a single random variable, X. The *DF* uses the *AM* and the *EM* of the N observations. The variance of the computed *DF* requires the derivation of the variances of the components of the formula. The variance of the *AM* is as follows,

$$\begin{aligned} \mathrm{V}(AM) &= (1/N) \times \mathrm{V}(X_1) + (1/N) \times \mathrm{V}\,(X_2) + \ldots + (1/N) \times \mathrm{V}(X_N) \\ &= N \times \mathrm{V}(X_i);\ i \in \{1, 2, \ldots, N\} \end{aligned} \tag{A7}$$

since the X_i are random variables. The variance of the *DF* is,

$$\begin{aligned} \mathrm{V}(DF) &= \mathrm{V}[(1/EM) \times (AM - EM)] \\ &= \mathrm{V}[(1/EM) \times AM)] \qquad [(\mathrm{V}(EM) = 0)] \\ &= (1/EM)^2 \times (\mathrm{V}(X)/N). \end{aligned} \tag{A8}$$

Substituting (A2) and (A6) into (A8) gives,

$$V(DF) = \frac{1}{[90/(N \times (10^{1/N} - 1))]^2} \times \frac{900 \times (N11r - N11 - 9r - 9)}{N^3 \times (r^3 - r^2 - r + 1)}$$
$$= \frac{N11r - N11 - 9r - 9}{9N \times (r + 1)}. \tag{A9}$$

The standard deviation (SD) is the square root of (A9). An approximation of the SD is

$$SD(DF) = .63825342/\sqrt{N}. \tag{A10}$$

The approximation (A10) was originally derived using integral calculus and was used in this study. Equation (A10) is used to formulate prediction intervals using the same logic used to construct confidence intervals in hypothesis testing. The computed *DF*s are restricted in range as follows:

$$DF_{min} = (AM_{min} - EM)/EM$$
$$\cong (10 - 39.08)/39.08$$
$$\cong -0.7441; \tag{A11}$$
$$DF_{max} = (AM_{max} - EM)/EM$$
$$\cong 1.5586, \tag{A12}$$

where *min* and *max* denote minimum and maximum values respectively. An *EM* of 39.08 is the approximate mean of a relatively large *collapsed* Benford set.

Since the *AM* is the mean of a set of random variables, the distribution of the *DF*s will, under the central limit theorem, tend towards the normal distribution for large *N*. Using the normal approximation, lower and upper bound prediction intervals (abbreviated PI(*lb*) and PI(*ub*)) for the determination of *DF* values such at $PI(lb \leq DF \leq ub) = 0.95$ are as follows:

$$PI(lb) = 0 - 1.96 \times SD(DF); \tag{A13}$$
$$PI(ub) = 0 + 1.96 \times SD(DF). \tag{A14}$$

APPENDIX B
PARTITIONING CRITERIA FOR OBSERVATIONS ON 1985 AND 1988 ITMF

Panel A: 1985: Observations classified as High Income (lower incentive)

1. If any of the following conditions were met

If filing period other than 1985; Filed outside of continental U.S.; Marginal tax rate equal to zero.

2. If any of the following exceeded $100,000

AGI; Schedule C, E or F net income; Net capital gain; Total itemized deductions; Current short-term gains; Current short-term losses; Residence gain; Current long-term gains; Current long-term losses; Rent net income; Royalty net income; Total partnership income; Total partnership loss; Total small business corporation income.

3. If any of the following were less than $0

AGI; Schedule C, E or F net income; Net capital gain; Rent net income; Royalty net income.

Panel B: 1988: Observations classified as High Income (lower incentive)

1. If any of the following conditions were met

If filing period other than 1988; Filed outside of continental U.S.; Marginal tax rate equal to zero.

2. If any of the following exceeded $100,000

AGI; Schedule C, E or F net income; Net capital gain; Excess itemized deductions; Schedule C receipts; Current short-term gains; Current short-term losses; Current long-term gains; Current long-term losses; Rental income; Royalty income; Rent net income; Royalty net income; Total passive income; Total non-passive income; Combined partnership and S corporation net income; Total passive losses; Total losses allowed.

3. If any of the following were less than $0

AGI; Schedule C, E or F net income; Net capital gain; Combined partnership and S corporation net income; Royalty net income; Rent net income; Total passive losses; Total non-passive losses.

REFERENCES

Alm, J., B. Jackson, and M. McKee. 1992. Estimating the determinants of taxpayer compliance with experimental data. *National Tax Journal* 45 (March): 107–114.

Benford, F. 1938. The law of anomalous numbers. *Proceedings of the American Philosophical Society* 78 (March): 551–572.

Carroll, J. 1989. A cognitive-process analysis of taxpayer compliance. In *Taxpayer compliance. Volume 2: Social science perspectives*, edited by J. Roth, and J. Scholtz. Philadelphia, PA: University of Pennsylvania Press.

Carslaw, C. 1988. Anomalies in income numbers: Evidence of goal oriented behavior. *The Accounting Review* 63 (April): 321–327.

Christian, C., and S. Gupta. 1993. New evidence on "secondary evasion." *The Journal of the American Taxation Association* 15 (Spring): 72–92.

Cowell, F. 1985. Tax evasion with labour income. *Journal of Public Economics* 26: 19–34.

Cuccia, A. 1994. The economics of tax compliance: What do we know and where do we go? *Journal of Accounting Literature* 13: 81–116.

Dusenbury, R. 1993. Evidence From a Tax Model File on the Effect of the Prepayment Position on Individual Taxpayers. Working paper, Florida State University, Tallahassee, FL.

Fischer, C., M. Wartick, and M. Mark. 1992. Detection probability and taxpayer compliance: A review of the literature. *Journal of Accounting Literature* 11: 1–46.

Hill, T. P. 1995. Base-invariance implies Benford's Law. *Proceedings of the American Mathematical Society* 123 (March): 887–895.

Jackson, B., and V. Milliron. 1986. Tax compliance research: Findings, problems, and prospects. *Journal of Accounting Literature* 5: 125–165.

Kidder, R., and C. McEwen. 1989. Taxpaying behavior in social context: A tentative typology of tax compliance and noncompliance. In *Taxpayer compliance. Volume 2: Social science perspectives*, edited by J. Roth, and J. Scholtz. Philadelphia, PA: University of Pennsylvania Press.

Klepper, S., and D. Nagin. 1989. The anatomy of tax evasion. *Journal of Law, Economics, and Organization* (Spring): 1–24.

Lemons, D. S. 1986. On the numbers of things and the distribution of first digits. *American Journal of Physics* 54 (September): 816–817.

Long, S., and J. Swingen. 1991. Taxpayer compliance: Setting new agendas for research. *Law & Society Review* 25 (3): 637–683.

Nigrini, M. J. 1994. Using digital frequencies to detect fraud. *The White Paper* (April/May): 3–6.

Pinkham, R. S. 1961. On the distribution of first significant digits. *Annals of Mathematical Statistics* 32: 1223–1230.

Raimi, R. A. 1969. The peculiar distribution of first digits. *Scientific American* 221 (December): 109–120.

———. 1976. The first digit problem. *American Mathematical Monthly* 83 (August-September): 521–538.

Rosch, E. 1975. Cognitive reference points. *Cognitive Psychology* 7 (October): 532–547.

Roth, J. A., J. T. Scholtz, and A. D. Witte, eds. 1989. *Taxpayer Compliance. Volume 1: An Agenda for Research.* Philadelphia, PA: University of Pennsylvania Press.

SAS Institute Inc. 1985. *SAS User's Guide: Statistics, Version 5 Edition.* Cary, NC: SAS Institute Inc.

Slemrod, J. 1985. An empirical test for tax evasion. *Review of Economics and Statistics* 69: 232–238.

The Wall Street Journal. 1995. Tax Report: Fix them or dump them! Those dreaded IRS super-audits draw heavy fire. (May 3).

Thomas, J. K. 1989. Unusual patterns in reported earnings. *The Accounting Review* 64 (October): 773–787.

U.S. Internal Revenue Service (IRS). 1985. *General Description Booklet for the 1985 Individual Public Use Tax File.* Washington, D.C.: U.S. Dept. of the Treasury, Statistics of Income division.

———. 1988. *General Description Booklet for the 1988 Individual Public Use Tax File.* Washington, D.C.: U.S. Dept. of the Treasury, Statistics of Income division.

———. 1990. *Income Tax Compliance Research: Net Tax Gap and Remittance Gap Estimates* (Supplement to Publication 7285). Washington, D.C.: U.S. Dept. of the Treasury.

———. 1991. *Internal Revenue Service: Highlights 1990* (Publication 1265, Rev. 7-91). Washington, D.C.: U.S. Dept. of the Treasury.