
Exploring Current and Potential Markets for Electronic Filing Using a Tree-Structured Analysis

Anne Steuer and Michelle Benson, Internal Revenue Service¹

Electronic filing (ELF) is an important component of the Internal Revenue Service (IRS) tax systems modernization effort. In 1986, the IRS initiated ELF as a test in three metropolitan areas, and, by 1990, the ELF program was available nationwide. Electronic filing is beneficial to both taxpayers and the government. With electronic filing, taxpayers receive refunds faster, omissions or errors on returns are essentially eliminated, and return information is accurately posted to taxpayer accounts while the government realizes reduced processing costs.

In the past, the IRS has focused much of its marketing efforts on expanding existing ELF market segments, particularly paid return preparers. By promoting ELF among professional tax preparers, the IRS can reach a broad segment of taxpayers. In Minnesota, for example, 64.5 percent of ELF eligible returns for Tax Year 1993 were prepared by professional tax return preparers. The paid preparer community has, in turn, emphasized ELF as a way to receive a relatively fast tax refund, often through a "refund anticipation loan." These relatively high-cost loans are utilized more frequently by those taxpayers most in need of their refunds--lower income, younger filers claiming the earned income credit (GAO, 1993; Musselman, 1991).

A 1995 GAO report on electronic filing stressed the need to explicitly consider potential cost savings to the government in expanding the ELF program. Higher cost savings are realized through electronic filing of the more complex forms and schedules (Forms 1040). Most of the ELF returns, almost 75 percent, could have been filed on the simpler forms--1040EZ and 1040A. This is before the availability of the TeleFile program, which is directed towards 1040EZ filers.

This paper examines current and potential market segments of individual taxpayers who are eligible to file electronically in the state of Minnesota using a tree-structured analysis.

■ Data

The data used for this analysis consist of individual taxpayers who filed ELF or were eligible to file ELF for Processing Year 1994 and filed their Federal 1040 returns from the State of Minnesota. The data were obtained from the Interim Compliance Research Information System (ICRIS). These data are a stratified random sample of taxpayers who filed returns in Processing Year 1994 (PY94).

Data on cost savings of ELF returns as compared with paper returns were obtained from a 1995 GAO report on electronic filing, North Central District Taxpayer Service, and Document 6746 (Rev. 1/93). Cost savings include upfront processing costs only, such as coding, editing, data entry, simple error correction, and mailing of refunds. Additional cost savings are realized from electronic filing, such as direct deposit of refunds, storage and retention costs, reduction in errors (by 90 percent over paper), archiving, etc. Although relevant, no current actual figures have been calculated to identify these amounts or related cost savings per return.

■ Methodology

Potential ELF market segments having the highest "cost savings" were identified using CHAID (Chi-squared Automatic Interaction Detector) version 6.0 with SPSS. CHAID is a segmentation modeling approach, which can be used to divide a population into segments (or groups) based on the values of a categorical dependent variable and a set of categorical explanatory variables. The process of partitioning the data is sequential and proceeds in a forward direction. The algorithm selects explanatory variables using statistical hypothesis testing based on an appropriate formal test statistic as described below. In this analysis, the significance level was set at 5 percent for all statistical tests. This significance level is adjusted for simultaneous inference using the Bonferroni multiplier.

For this application, the dependent variable has five categories based on form type filed--current ELF filers, potential 1040PC filers, potential 1040EZ filers, potential 1040A filers, and potential 1040 filers. 1040PC forms are prepared on a PC using appropriate software, printed, and filed as paper returns. 1040EZ filers have the least complex returns. Main requirements include filing status of single or married filing jointly, claim no dependents, taxpayer age 65 years or less, only wage income with \$400 or less taxable interest income, total taxable income of \$50,000 or less, and no itemized deductions or tax credits. 1040A filers meet the above 1040EZ profile with the following additional items allowed: all filing statuses, Social Security benefits, IRA payments, unemployment compensation, and tax credits--earned income credit, credit for the elderly, and child care credit. 1040 filers can fit either of the 1040EZ or 1040A profiles. However, this form is normally required when there are any unusual circumstances or thresholds are exceeded in any area for the 1040EZ or 1040A profiles. Each of these form types has a different potential cost savings comparing the paper version to the electronically filed version. The ordinal character of the dependent variable is captured through a vector of scores, cost-savings information in this case.

Initially, the data are partitioned into groups based on the values of the most statistically significant explanatory variable under the following hypothesis test:

For an ordinal dependent variable, Y , with J categories and a set of monotonically increasing (or decreasing) scores $\{V_j\}$ and any nominal categorical explanatory variable, X , with I categories, expected cell counts, F_{ij} , are described by the following loglinear model for ordinal-nominal two-way contingency tables (Agresti, 1984; Magidson/SPSS, Inc., 1993):

$$\ln\left(\frac{F_{ij}}{Z_{ij}}\right) = \lambda + \lambda_i^X + \lambda_j^Y + \tau_i(v_j - \bar{v})$$

where λ_i^X is the i^{th} row effect, λ_j^Y is the j^{th} column effect, and $\tau_i(v_j - \bar{v})$ is the effect due to the association between X and Y . In addition, $\sum_i \lambda_i^X = \sum_j \lambda_j^Y = \sum_i \tau_i = 0$ and $\{v_j\}$ are known. Z_{ij} , the reciprocal of the average cell weight, is an adjustment for weighted data (Clogg and Eliason, 1987). The independence model

corresponds to the case where the parameters $\tau_1 = \tau_2 = \dots = \tau_I$ so that

the hypothesis test takes the following form:

$$H_0: \tau_1 = \tau_2 = \dots = \tau_I$$

$$H_1: \ln\left(\frac{F_{ij}}{Z_{ij}}\right) = \lambda + \lambda_i^X + \lambda_j^Y + \tau_i(v_j - \bar{v})$$

The likelihood ratio test statistic for this test is

$$2 \sum_i \sum_j \hat{F}_{ij}^{H_1} \ln\left(\frac{\hat{F}_{ij}^{H_1}}{\hat{F}_{ij}^{H_0}}\right)$$

which is distributed approximately X^2 with $I - 1$ degrees of freedom. Here, $\hat{F}_{ij}^{H_0}$ and $\hat{F}_{ij}^{H_1}$ are the estimated expected cell counts under the null and alternative hypotheses, respectively.

After an explanatory variable is selected, pairs of categories of the variable are considered for merging based on the above test using a 2 by J sub-table of the explanatory variable category pair and the dependent variable. If the test statistic for the 2 by J sub-table is not statistically significant, the two categories are merged. After this process, each category or all categories of the explanatory variable form the "branches" or "nodes" for the next set of splits (or groupings).

The algorithm selects the next explanatory variable for each branch using the most statistically significant variable from among the remaining explanatory variables. The process continues until there are no more statistically significant explanatory variables or a user-supplied stopping rule is encountered.

The cost savings associated with each of these form types (relative to the status quo--it was assumed that ELF filers would continue to file ELF returns) was estimated to be \$0.00 for current ELF, \$0.66 for 1040PC, \$0.28 for 1040EZ, \$0.87 for 1040A, and \$1.45 for 1040 filers. These costs represent the upfront processing costs, including opening and sorting mail, transcribing data, and upfront validity checks.

Predictor variables used in the analysis include:

Adjusted Gross Income (AGI)--Categories are: less than \$13,000; \$13,000 to \$26,000; \$26,000 to \$39,000; \$39,000 to \$52,000; and greater than \$52,000.

Age--Age of taxpayer for nonjoint returns; age of primary taxpayer for joint returns. Categories are: less than 16; 16 to 20; 21 to 24; 25 to 44; 45 to 64; greater than 65; and unknown.

Refund/Balance Due--Categories are: greater than \$3,000 refund; \$1,000 to \$3,000 refund; \$300 to \$1,000 refund; \$1 to \$300 refund; \$0 balance; \$1 to \$300 due; \$300 to \$1,000 due; \$1,000 to \$3,000 due; and greater than \$3,000 due.

Tax Credits--Categories are: no credits; credit for the elderly (CFE) only, earned income credit (EIC) only, child care credit (CCC) only; EIC and CCC only; EIC and CFE only; CCC and CFE only; EIC, CCC, and CFE.

Paid Preparer Indicator--Categories are: use of paid preparer or self-prepared return.

Overpayment/Balance Due Indicator--Categories are: return with overpayment; return with no overpayment or balance due; return with balance due, fully paid; return with balance due after remittance.

Geographical Location--Categories are: Minneapolis/St. Paul; Duluth; St. Cloud; Rochester; Mankato; rural and/or out of state.

The refund/balance due categories were constructed to capture information about refund anticipation loans. These loans typically range from \$300 to \$3,000. Age categories follow standard industry groupings. Adjusted gross income (AGI) categories were constructed to obtain information on earned income credit (EIC) eligibles. EIC was at a maximum at \$13,000 and phased out at \$26,000 in 1993. AGI categories also follow closely with tax rate divisions for common filing statuses.

■ Results and Discussion

There were an estimated 2,005,127 (98.8 percent) current and potential ELF filers out of 2,030,236 filers in Minnesota for Processing Year 1994. Of those, an estimated 176,043, or 8.78 percent, are ELF filers. Most market segments of ELF filers include the use of a paid preparer. Overall, an estimated 142,570, or 81 percent, of current ELF filers used a professional tax preparer.

This result is expected as most ELF filing must be done through a professional tax return preparer or other business that will transmit self-prepared returns, usually for a fee. ELF is available at no cost to low income and elderly taxpayers through Voluntary Income Tax Assistance (VITA) and other volunteer return preparer sites, although this accounts for less than 2 percent of the total volume of returns in Minnesota. Table 1 lists market segments with the highest concentration of current ELF filers in Minnesota. These market segments account for roughly half of all ELF filers for Tax Year 1993.

The estimated percentage of current ELF filers in these market segments ranges from 19.39 percent to 66.07 percent and the estimated market size from 2,127 to 235,177. Two of these market segments are broken down into smaller subgroups. The segment with the highest percentage of current ELF filers--tax refunds between \$300-\$3,000; taxpayers 16-24 years old utilizing a professional return preparer--has been subdivided into two groups--returns with and without tax credits. Although the group with tax credits has the highest concentration of ELF filers, it represents only a quarter of the taxpayers in this market segment. This group is made up primarily of young low income taxpayers claiming EIC and CCC. In Tax Year 1993, the EIC was available only to taxpayers with a qualifying child. In later tax years, EIC is also available to taxpayers without a qualifying child, provided the taxpayer is between the ages of 25-64 years and meets the income criteria for the tax credit.

The market segment--\$300-\$3,000 tax refund; < 45 years old; AGI > \$39,000; use of a paid preparer--was also subdivided into two groups based on the size of returns. The group with the larger refunds--\$1,000 to \$3,000--has a larger percentage of ELF filers (24.52 percent), compared with the lower refund group--\$300-\$1,000 (19.39 percent). While the taxpayers in this segment are not low income, they are relatively young as a group and may have financial pressures increasing their demand for a fast tax refund.

The largest market segment (235,177 estimated taxpayers) of high proportion current ELF filers from Table 1--\$300-\$3,000 tax refund; 25-64 years old; AGI <

Table 1. Market Segments of High Proportion Current ELF Filers.

Market Segment	% Current ELF Filers*	Estimated Size**	Average Cost Savings	Estimated Total Cost Savings
\$300 - \$3,000 tax refund; 16 - 24 years old; use of paid preparer	46.08	31,839	\$0.59	\$18,785
Of these returns . . .				
Returns with tax credits present	66.07	8,161	\$0.40	\$3,264
Returns with no tax credits	39.20	23,678	\$0.66	\$15,627
\$1 - \$300 tax balance due; balance due after remittance; use of paid preparer	45.39	12,439	\$0.66	\$8,210
Tax refund > \$3,000; earned income credit with or without other credits	30.42	2,127	\$0.94	\$1,999
\$300 - \$3,000 refund; 25 - 64 years old; AGI < \$26,000	22.36	235,177	\$0.80	\$188,142
\$300 - \$3,000 tax refund; < 45 years old; AGI > \$39,000; use of paid preparer	22.22	72,440	\$1.09	\$78,960
Of these returns . . .				
Returns with \$1,000 - \$3,000 tax refund	24.52	39,931	\$1.06	\$42,327
Returns with \$300 - \$1,000 tax refund	19.39	32,509	\$1.13	\$36,735
\$1 - \$300 tax refund; no tax credits; < 25 years old; use of paid preparer	16.61	46,279	\$0.87	\$40,263

* Percentage of current ELF filers in the particular market segment.
 ** Estimated size of the particular market segment of potential ELF filers.

\$26,000--consists of lower income taxpayers.

These results support the contention that electronic filing participation is concentrated mainly among groups of taxpayers most in need of their refunds (Musselman, 1991).

The market segment--\$1-\$300 tax balance due; balance due after remittance; use of a paid preparer--was an unexpected finding. These taxpayers filed their returns through a professional preparer, obtaining a pre-printed voucher (Form 1040V) to be submitted with payment before the return due date. This arrangement can be advantageous to the taxpayer as the return is received, accepted, and processed early, and any errors and omissions can be corrected before the due date possibly avoiding penalties and interest. In Minnesota, many professional preparers offer ELF at no additional charge to their customers. Some seasonal occupations, such as farming and fishing, are particularly suited to this arrangement of early filing with a balance due after remittance and later payment before the due date.

The market segment--\$1-\$300 tax refund; no tax credits; < 25 years old; use of a paid preparer--is apparently the result of ELF marketing efforts in Minnesota

directed toward students on college campuses.

Given the size of many of these segments and the current percentages of ELF filers, there appears to be potential for increasing the numbers of ELF filers in these segments. However, if cost savings to the government are explicitly considered, the most "cost-saving" segments appear to be quite different. Table 2 lists the principal potential cost-savings market segments, together with the average cost savings per return and the estimated market size. Many of these markets involve taxpayers with a balance due, which is fully paid with the return. The two market segments of taxpayers with refunds due with relatively high potential cost savings do not typically use paid preparers.

The average potential cost savings for these segments range from \$1.31 to \$1.42 per return. By contrast, the average potential cost savings for the major segments of current ELF filers presented above range from \$0.40 to \$1.13 per return. If geographical location (rural versus city) is considered, these cost figures can often be improved by an additional \$0.10 to \$0.20. These additional savings are due primarily to the differences in the mix of form type filers in a rural versus urban set-

Table 2. Principle ELF Filer Market Segments with the Greatest Potential Cost Savings.

Market Segment	Average Cost Savings	Estimated Size**	Estimated Total Cost Savings	% Current ELF Filers*
\$1 - \$300 tax refund; no tax credits; > 24 years old; AGI > \$52,000; no paid preparer	\$1.42	12,142	\$17,242	0.31
\$1 - \$300 tax due; fully paid with return; no tax credits; > 24 years old; AGI > \$52,000	\$1.40	21,878	\$30,629	0.06
No balance due on return; AGI > \$13,000	\$1.39	51,279	\$71,278	1.13
Tax balance due > \$300; fully paid with return	\$1.36	296,483	\$403,217	0.10
\$300 - \$3,000 tax refund; AGI > \$39,000; no paid preparer	\$1.32	100,968	\$133,278	4.07
No balance due on return; AGI < \$13,000; > 44 years old	\$1.31	41,721	\$54,655	1.82
\$1 - \$300 tax due; fully paid with return; tax credits claimed	\$1.31	13,874	\$18,175	0.23

* Percentage of current ELF filers in the particular market segment.

** Estimated size of the particular market segment of potential ELF filers.

ting. In the rural areas, there is a larger concentration of Form 1040 filers and a smaller concentration of 1040A filers, especially at the lower incomes, compared with urban areas. The smaller inventory of rental housing, and generally lower real estate values, may result in a larger concentration of real property owners. These taxpayers would typically file a Schedule A precluding the use of Form 1040A. Additionally, in rural Minnesota, there are fairly large concentrations of 1040 taxpayers filing a Schedule F (farm income).

The two market segments with the greatest potential average cost savings (\$1.40 - \$1.42) involve higher income taxpayers (AGI > \$52,000) who either owe tax, which is fully paid with the return, or receive a small refund. The largest market segment (estimated 296,483 taxpayers) from Table 2 consists of taxpayers with a tax balance due > \$300, which is fully paid with their returns. These groups have little incentive to pay for ELF service when they owe money or have no need for an immediate refund. With cost being the single largest deterrent to electronic filing, these results present a challenge to some of the traditional ELF marketing strategies focusing on the use of paid preparers and the em-

phasis on a fast refund.

Musselman (1991) has noted that taxpayers appear to be sensitive to the cost of electronic filing. In a survey of taxpayers, of those taxpayers stating that they would "definitely use" or "probably use" electronic filing, less than one quarter would be willing to pay more than \$20 for the service, but about one half would be willing to pay between \$10 and \$20 (assuming filing ELF would result in a refund within five days). Even if the cost of ELF can be reduced, one of the primary marketing tools--emphasis on a fast refund--may be of limited use as most of the greatest potential cost savings market segments consist of taxpayers with no refunds.

An interesting observation is that some cost-savings segments do not explicitly involve the use of a paid preparer. The IRS historical marketing focus on professional preparers may overlook these segments.

If the IRS wants to increase use of the ELF program, then it may want to consider offering alternatives to the segment who currently prepare their own returns and have small or no refunds (refunds less than \$300).

This group, historically, sees little benefit to paying for the service of electronic filing. An accepted return that has been validated for errors does not alone provide a sufficient incentive to this group of filers. With increases in technology, the IRS may want to consider expanding on the partnership with software and transmission companies to allow customers to file from home computers or kiosk-type centers. Low or no cost and convenience are still critical to entice these filers to participate (GAO, 1993).

Marketing has been limited to 1040 instruction inserts and businesses offering electronic filing services in the past. The IRS needs to continue to reach the public by using new methods and updating information to potential ELF filers, both the businesses offering this service and the public who will utilize it but do not currently participate. Many filers still base decisions on outdated information and from sources other than the IRS. Focus group research conducted by the IRS in the early 1990's indicated that the public was often not aware of ELF or believed the cost to the taxpayer was excessive (GAO, 1993). Messages from sources other than the IRS are often self-serving and do not necessarily present a complete or accurate reflection of the current ELF program.

■ Summary

Analysis of current and potential ELF filers has generally confirmed anecdotal information about the location and size of market segments of current ELF filers. Incorporating cost-savings information into the segmentation analysis has allowed potential and current ELF segments to be prioritized by average cost savings in addition to their estimated sizes. Analysis of these segments has identified some segments with relatively large potential average cost savings that may not be reached under current marketing strategies. If the IRS wants to expand participation in electronic filing, it may need to continue to expand options in the ELF program to entice more filers from market segments not typically using this option.

■ Acknowledgements

The authors would like to thank Ron Bartyczak, Brian Musselman, Michael Gregory, and Donald A. Berry for their review and suggestions.

■ Footnote

- ¹ The views expressed in this article represent the opinions and conclusions of the authors. Such views do not necessarily represent the position of the Internal Revenue Service.

■ References

- Agresti, A. (1984), *Analysis of Ordinal Categorical Data*, John Wiley and Sons, Inc., New York.
- General Accounting Office, January 1993, Opportunities to Increase the Use of Electronic Filing, GAO/GGD-93-40.
- General Accounting Office, October 1995, Electronic Filing Falling Short of Expectations, GAO/GGD-96-12.
- Clogg, C. and Eliason, S. (1987), Some Common Problems in Log-Linear Analysis, *Sociological Methods and Research* 16(1):8-43.
- Internal Revenue Service, Ogden Service Center Document 6746 (Rev. 1/93), Costing Reference for Service Center Processing, OSC Public Affairs, April 1993.
- Magidson, J. (1992), Chi-Squared Analysis of a Scaleable Dependent Variable, *Proceedings of the Section on Statistical Education of the American Statistical Association*, pp. 242-247.
- Magidson, J. (1993), The Use of the New Ordinal Algorithm in CHAID to Target Profitable Segments, *The Journal of Database Marketing*, 1(1):29-48.
- Magidson, J. and SPSS, Inc. (1993), SPSS for Windows CHAID Release 6.0 Users Manual, SPSS, Inc., Chicago, IL, 145 p.
- Musselman, B. L. (1991), Electronic Filing--Who's Participating and Who Isn't, *The IRS Research Bulletin*, U.S. Department of the Treasury, Internal Revenue Service, Publication 1500, pp. 27-32.