

## **EITC Take-up**

*A look at Earned Income Tax Credit take-up rates by low- to moderate-income individuals, as determined by tax preparation firms' locations and services.*

By

Jacque Dodd, Lisa Poplawski & Peter Zochowski

Presented to the Department of Economics, University of Oregon,  
in partial fulfillment of requirements for honors in Economics.

Under supervision of  
Prof. Bill Harbaugh

Spring 2008

## EITC Take-up

*A look at Earned Income Tax Credit take-up rates by low- to moderate-income individuals, as determined by tax preparation firms' locations and services.*

**Abstract:** Using data on tax preparation firms, tax payer characteristics and U.S. Census we determine correlation between Earned Income Tax Credit take-up rates and number of tax preparation firms in a ZIP code. Results indicate a firm's decision to locate as based on high EITC take-up rates, as well as EITC take-up rates influenced by additional firms in the area. Analysis suggests difficulty for volunteer tax preparation services such as United Way of Lane County to compete with for-profit firms, because of Refund Anticipation Loan options for-profit firms are exclusively able to offer low-income customers; a service valued by this demographic as a means for immediate credit funds.

*Keywords: Earned Income Tax Credit; EITC; United Way of Lane County; Volunteer tax preparation services; Refund Anticipation Loans; RALS*

**Approved:** \_\_\_\_\_

Prof. Bill Harbaugh

Date

## Table of Contents

Introduction	4
Literature Review	7
Hypothesis Development & Model Specification	12
Data Description	14
Firm Location (IRS)	15
Taxpayer Characteristics (IRS-SPEC)	16
Census (U.S. Census Bureau)	18
Methodology & Empirical Results	18
Conclusion	24
Recommendations	28
Appendices	33

## **Introduction**

Since its instatement in 1975, the Earned Income Tax Credit (EITC) has helped bring billions of dollars in assistance to low- and moderate-income (LMI) working families. A refundable federal income tax credit, EITC increases its recipient's income by providing a credit calculated by income and family size. In 2004 alone, EITC credits provided approximately \$39 billion to more than 21 million low-income families. Despite the large financial benefit associated with the EITC, 15-20 percent of eligible workers fail to claim their EITC credit. The largest explanation for this missed percentage is failure to file a tax return. This deficit translates into almost \$11 billion in unclaimed credit (Welfare Peer, 2006).

Community volunteer organizations have started many outreach programs through the Internal Revenue Service's (IRS) Volunteer Income Tax Assistance (VITA) program to help LMI families claim their EITC credit. However, for-profit tax preparation firms are increasingly realizing the revenue potential of the EITC refund market. Furthermore, a large portion of individuals who are EITC eligible have no formal relationship with any banking institution. A push by these firms to target LMI individuals has caused controversy surrounding over-charging and specifically targeting LMI customers because of their typical characteristics—one being hyperbolic discounting and thus increasing their likelihood to agree to charges and fees for immediate money in the form of Refund Anticipation Loans (RALs) (Welfare Peer 2006).

The profitability for all tax preparation firms to bring financial and tax preparation services to the LMI customer has driven firms like H&R Block to locate at almost every corner, raising their total for the 2007 company-owned office count to 9,036 (H&R Block Press Center). The cause for concern here is the diversion of EITC tax credit refunds from low-income families to tax preparation firms in the forms of fees and RAL interest. The Welfare Peer Technical

Assistance Network, under the U.S. Department of Health and Human Services, reports that in 2004, at least half of all EITC returns were prepared by paid firms. While large firms like Jackson Hewitt, Liberty Tax Services and H&R Block have stopped charging customers “administrative” and “application” fees, the average cost per customer is still at \$300, reducing the average EITC per person credit by 17 percent, from \$1750 to \$1450 (Welfare Peer, 2006).

Locally, United Way of Lane County, a nonprofit that serves over 100,000 Lane County men, women and children in human service areas, offers volunteer tax preparation programs for all community members and hopes to attract LMI families who need tax assistance and guidance. United Way’s efforts attempt to deliver EITC credit to all eligible community members, while at the same time diminish fees and interest on loans that tax preparation firms charge. For United Way to reach its goal of allowing more eligible families to take advantage of the EITC without penalizing them in fees, United Way competes with for-profit tax preparation firms who have both the funds and associates to constantly build up firm count and increase advertising of services. Most importantly, however, the tax preparation firms are able to offer eligible clients RALs, which allows them to deliver EITC refunds immediately at the price of the loan interest rate and preparation fee. With most LMI individuals characterized by hyperbolic discounting and therefore an immediate demand for their EITC, United Way will lose in this market if demand for accessibility and immediacy of tax preparation services overrides the lost utility in fees these firms charge. Motivated by United Way’s local inquiry into delivering EITC to as many eligible community members as possible, this project explores the reality of competing with for-profit firms, and determines the benefit, or lack thereof, of United Way spending their time and resources on this service.

With little data available from United Way, the fundamental understanding of EITC take-up rates and tax preparation firm location decisions must be explored, and make up the bulk of this project. Determining a firm's decision to locate and common EITC demographic characteristics will allow an understanding of United Way's position in the tax preparation market. The focus of this study, then, is to determine what influences EITC take-up rates, given the number of tax preparation firms in a region. For our analysis, tax firms Jackson Hewitt, H&R Black, Liberty Tax Services and various payday lender sites will be used since they offer RALs.

In 2004, 96 percent of EITC recipients had incomes below \$30,000, and 71 percent had incomes below \$20,000 (Welfare Peer, 2006). Because the demographic for EITC recipients is easily identifiable, our regression model will include census indicators for this segment of the population. This allows us to verify the motives behind tax preparation firm placement as a means of reaching more low-income customers and therefore generating more RALs and tax preparation fees. Necessary components of our project include research on and discussion of relevant groups including the unbanked, tax preparation firms, and EITC recipient characteristics, as well as background research on EITC both in theory and in practice.

Four models are considered for this project. In each of our regressions, we will include several demographic dependent variables from our census datasets. Our independent variables will be the number of firms, EITC take-up, and number of RALs. From these models we will be able to determine the marginal effects of each of these variable changes, while holding constant the demographic characteristics in each ZIP code. We expect that more firms will result in increased take-up rates and more refunds through RALs. In addition, we expect for-profit firms to locate in areas that already have a high rate of EITC take-up.

## **Literature Review**

To fully understand the EITC, we must be aware of its historic changes and the economic results of each instituted adjustment. This allows us to gauge the effectiveness of volunteer tax assistance programs and potential strategies to increase their impact. Since its implementation, there have been numerous revisions and expansions with the goal of reducing payments to ineligible EITC filers without causing harm to, or deterring the participation of, eligible low-income working families. The ultimate objective is to eliminate benefits to ineligibles while awarding credits to all eligible citizens. Modifications to the program include the expansion of EITC payments and the number of eligible taxpayers, the simplification of rules, along with the addition of required forms and documents.

The Omnibus Budget Reconciliation Act of 1993 (OBRA93) was implemented as an expansion of EITC payments, eligible taxpayers and required tax return forms. Once phased in, the OBRA93 reforms were estimated to substantially increase the number of individuals receiving the credit, the amount of Federal funding allocated toward closing the poverty gap, and the measure of taxpayers whose incomes will be raised above the poverty line (Scholz, 1994). The OBRA93 reforms added a two-page form to deter false claims, in response to the empirical finding that increasing the financial benefits of the EITC increases the incentive for ineligible persons to falsely file (many times erroneously receiving payments). Changes such as these have been successful in extending the reach of Federal assistance to low-income workers, yet still fall short of the ambitious goals of the EITC.

Volunteer agencies providing free tax preparation services have been somewhat successful in combating the shortcomings of the EITC, yet various alterations instigated by the IRS have created further difficulties for such programs. As the IRS has attempted to abolish false

EITC claims by increasing the number of forms and information required to receive payment—which has enlarged the demand for free tax preparation services—the department has also reduced support provided through Taxpayer Assistance Centers. This has left under-staffed volunteer sites unable to meet the increased demand of consumers (Greenstein, 2003). A crucial recommendation to prevent further failure of the market for volunteer tax assistance is that the government provide a grant program for free tax preparation services to low-income taxpayers, and that volunteer agencies focus more resources aimed at acquiring the necessary technology for the current tax system.

By building onto this research with examination of the increasing dependence on paid tax preparation and entrance of low-cost tax assistance firms into the market, we can examine the ideal arrangement of the EITC and role of volunteer organizations. Taking into account the regression models and techniques used in previous studies, such as probit regression on eligible non-filers and demographic dummy variables used to proxy for lack of information, we will be able to assess an efficient and reliable way to obtain supported empirical findings.

Berube and the Brookings Institute (2006) released findings from an EITC analysis study that looked at changes in EITC's influence and monetary data from 2000 and 2003. This study shows changes in EITC recipients and credits, and highlights policies that increase EITC take-up and access. Most importantly, this study examines EITC refunds prepared by for-profit tax firms, allowing us to empirically analyze if these businesses are attempting to target LMI customers between 2000 and 2003. Berube used data from the IRS-Stakeholder Partnerships, Education, and Communication (SPEC) Return Information Database. This database contains information from the IRS Electronic Tax Administration Marketing Database, and groups tax returns by U.S. ZIP code and what are referred to as market segments—a division that breaks up tax return data

by EITC received and low-income filers. Both of these characteristics are relevant to our project and provide a helpful example of an analysis by demographic.

In addition to their 2006 report, Berube and the Brookings Institute conducted two previous studies on EITC. The biggest shortcoming of their reports is they do not provide the econometric models used, only the general statistics recorded. However, it appears the Brookings Institute has proprietary access to IRS data that is unavailable to the general public. Most notably, they have data on the location and name of the tax preparation office that filed the EITC and whether the individual receiving the EITC opted to use a RAL. Fortunately, Berube supplied us with these datasets for our own examination.

In his 2002 study, Berube determined that the spatial distribution of H&R Block and Jackson Hewitt offices is positively correlated to the concentration of EITC qualifiers in the area. In ZIP codes with large concentrations of individuals who qualify for the EITC, there are typically 50 percent more tax preparation firms, with this trend most prevalent in Southern and Western states. It was determined that Jackson Hewitt offices are seven times more likely to locate in ZIP codes with high concentrations of EITC qualifiers, in comparison to ZIPs with a only a large poor population that is not receiving the EITC. In areas where less than 10 percent of the population receives the EITC there are typically around 10 electronic return originators (ERO) —tax firms who have registered with the IRS to electronically-file tax returns. For every 10,000 filers, a five percent increase in the number of individuals who qualify for the EITC yields an increase of one ERO. We possibly can conclude that the EITC take-up rate increases as more professional tax-filing offices become available. Thus, it is beneficial for the public to have access to more tax preparation firms. However, it appears tax preparation firms have increasingly decided to build new offices in low-income areas to tap into the EITC market. This is supported

by the fact that during the past 20 years, Jackson Hewitt has increased from 15 to 3,300 offices, typically building in low-income areas. The EITC related payoff for tax preparation firms is big; nearly \$1.75 billion was spent on tax preparation in 1999 by tax payers filing for the EITC. RALs, which are relatively risk-free to lenders, provide a substantial portion of that profit. In 1999, 39 percent of all EITC refunds were distributed through RALs and, more importantly, nearly 50 percent of all money issued through the EITC was received through the RAL system. This suggests that the greater the EITC refund, the higher the likelihood the filer will opt for a RAL: Individuals increase their hyperbolic discount rate as the magnitude of the payoff increases (Berube, 2002).

The Federal Reserve figures that 22 percent of families earning less than \$25,000 a year are unbanked. A major hurdle for individuals collecting EITC is that many do not have a bank account to directly deposit their refund into. Given that many of these people are hyperbolic discounters, this increased waiting period makes RALs even more attractive. ShoreBank, a chain of banks located in Chicago, addressed this issue by offering low-cost bank accounts and free tax preparation services. Since this program began, they have helped 200 people directly deposit their EITC refund into a savings account. Of those 200 individuals, nearly 60 percent were previously unbanked. From this example, we see that large infusions of capital could motivate financial intuitions to provide services to the unbanked (Berube, 2001).

Much of the previous academic research on the unbanked has focused on identifying characteristics of this segment of the population, including lifestyle and race. For our purposes, race or minority status will not be a concern, but understanding statistics and lifestyles of the unbanked will help us relate EITC take-up possibilities for this population, as well as predict their behavior with tax preparation firms. An article published by J. Caskey (1994) discusses the

demographics of those who use banking services versus those who do not. Caskey reports the results of a national survey he conducted of 2,563 households in both 1977 and 1989, leading to results on trends of bank account ownership, and, more specifically, the differences between owning a checking versus a savings account. Caskey concludes that no one characteristic such as income, determines whether an individual will be unbanked or not, but instead that the 1977-89 decline in the number of households using financial services was due to socioeconomic changes that reduced financial wealth and, therefore, the need for financial services. Further and more recent academic research on the unbanked builds off Caskey's findings and survey techniques. In their working paper, Vermilyea and Wilcox (2002) use data from a household survey similar to Caskey's, administered between October 1998 and March 1999. This survey targeted populations that had a high likelihood of being unbanked, and aimed at grasping LMI individuals' attitudes toward banking and their previous banking activities. Vermilyea and Wilcox conclude, like Caskey, that socioeconomic characteristics affect choices about using a banking institution, and then, about which services to use within that institution. Both Caskey, Vermilyea and Wilcox's research methods and conclusions help us identify socioeconomic indicators of the unbanked to be used in our regression model, and provide us with general statistics on the unbanked to be used for our analysis.

In their working paper, Greene, Rhine and Toussaint-Comeau (2003) study the relationship between the unbanked population and check-cashing businesses. While their paper focuses on racial and ethnic differences, it does provide useful discussion on alternative ways the unbanked population carries out its financial transactions. Like other literature on the unbanked, Green et. al. uses survey data to answer characteristic questions on this population. Also like other literature on this topic, this paper is helpful with general discussion of the 15-20 percent

that is eligible, but does not access their EITC credit. Gaining information on their financial transactions with some type of firm, however, helps us gauge their ability to be reached and therefore eventually claim their EITC.

Finally, finding academic research on EITC theory gives us a good introduction into EITC benefits and initial intentions. Holtzblatt, McCubbin and Gillette (1994) examine EITC's incentive for work, as the tax-credit program requires its recipients to work, and increases—rather than decreases—with each dollar earned. Holtzblatt et. al. explain and analyze the effects of each EITC expansion through graphs and statistics of take-up since its beginning in 1975. By understanding EITC's foundation, this article suggests the harm tax preparation firms may be causing if the firms undercut a significant amount of the customer's EITC credit each tax year.

### **Hypothesis Development and Model Specification**

Our inability to locate time-series data on tax preparer locations was the biggest limiting factor. Had we used time-series variation for tax firms we would have been able to pin down the exact outcome an additional firm has on the EITC. The SPEC data was from the years 2000 and 2005, which gave us the option of using panel data regressions (STATA xtreg command). From the onset, we were interested in documenting what influences a firm's decision to locate and what characteristics determine the EITC take-up rate. To accomplish this, we picked several demographic variables that would be used in all regressions to define the attributes of a given ZIP code. These variables include: total population, marginal household income, number below the poverty line, education, average household size and number of non-white population. The theory behind using the same descriptive characteristics in every regression is to establish a baseline from which we could find the marginal effects of other changes. Therefore, with all

demographic variables held constant we can pinpoint the effect one more tax preparation firm has on the EITC take-up rate.

A possible problem with these demographic variables is that many are highly correlated to EITC eligibility, such as education, median household income and number under the poverty line. Thus, when we introduce variables relating to the number of people who received EITC, it is possible some of these income related variables become insignificant. For a firm's decision to locate, we hypothesize there to be a strong positive correlation between the total population, poverty line and the number who received EITC in the past. We believe that a major factor in a firm's decision to locate in a particular ZIP is to tap into the RALs market, which is correlated to the number of EITC recipients. Similarly, we hypothesize that the number of firms in a particular ZIP and the number of people under the poverty line will explain the EITC take-up rate.

$$\text{Firms Decision to Locate} = \beta_1 + \beta_2\text{TP} + \beta_3\text{MHI} + \beta_4\text{PL} + \beta_5\text{EITC} + \beta_6\text{E} + \beta_7\text{NW} + \mu$$

$$\text{Number of People Who Received the EITC} = \beta_1 + \beta_2\text{TP} + \beta_3\text{MHI} + \beta_4\text{PL} + \beta_6\text{E} + \beta_7\text{NW} + \beta_8\text{AHS} + \beta_9\text{RAL} + \mu$$

Where:

- TP= Total Population
- EITC = Total Returns with EITC
- MHI = Median Household Income
- PL= Number of People Below the Poverty Line
- E = Number of People with a High School Diploma or Better
- NW = Number of Non-White Persons
- AHS = Average Household Size
- RAL = Number of Firms That Offer RALs.

As we began to dig deeper into our data, we realized several other regressions that could explain the EITC. We decided to continue using the same basic demographic variables and check for robustness by adding additional variables. The first regression below identifies what characteristics influence a person's decision to take out a RAL. We theorize that simply having the option to take out a loan will increase the number of people who receive a RAL, thus, the number of RALs should be highly correlated with the number of firms. In addition, we want to see the effects of having a volunteer site in a ZIP code and determine whether consumers substitute the services of a paid provider for those of a volunteer organization.

$$\text{EITC Distributed With RAL} = \beta_1 + \beta_2\text{PL} + \beta_3\text{TP} + \beta_4\text{MHI} + \beta_6\text{NW} + \beta_7\text{E} + \beta_8\text{AHS} + \beta_9\text{RAL} + \mu$$

$$\text{Potential EITC Recipient} = \beta_1 + \beta_2\text{PL} + \beta_3\text{TP} + \beta_4\text{MHI} + \beta_6\text{NW} + \beta_7\text{E} + \beta_8\text{AHS} + \beta_9\text{RAL} + \beta_{10}\text{V} + \mu$$

Where:

- TP= Total Population
- EITC = Total Returns with EITC
- MHI = Median Household Income
- PL= Number of People Below the Poverty Line
- E = Number of People with a High School Diploma or Better
- NW = Number of Non-White Persons
- AHS = Average Household Size
- RAL = Number of Firms That Offer RALs.
- V = Number of Volunteer Sites

## Data Description

Our data can be divided into three categories; IRS firm location data, IRS SPEC (Stakeholder Partnerships, Education, and Communication) data on tax preparer characteristics,

and demographic data from the U.S. Census Bureau. See Appendix 1 for SPEC tax payer characteristics and Census summary statistics.

### Firm Location

Since the majority of EITC payments are distributed through paid tax preparers, it is critical to discover the marginal effects on the take-up rate of having an additional tax firm in a given ZIP code. In addition, we are curious to gain insight into common demographic characteristics that are in ZIP codes in which firms are located. The IRS website proved invaluable for data on the distribution of firms. When filing taxes, a taxpayer has two options, electronically file (E-file), or send in a paper copy via the mail. It has been well-documented that over 50 percent of EITC refunds are filed electronically, with H&R Block and Jackson Hewitt responsible for the vast majority of these filings (Berube, 2000). In order for a firm to offer E-filing services they must fill out an application with the IRS, which includes common attributes such as business name, address, phone number etc. These firms are then deemed Electronic Return Originators (ERO). Information on the location of particular EROs allows us to determine common characteristics that influence a tax preparation firm to locate in a given area. Once filed with the IRS, this information becomes public and is accessible via the IRS website. We decided it would be best to include only data on firms that offer RALs, as their ability to offer immediate EITC funds distinguishes them from volunteer sites like United Way, and is a service that predictably draws many LMI customers. Distinguishing by RALs offerings divides the firms into CPA offices and franchised tax preparation firms (Jackson Hewitt, Liberty Tax Services, H&R Block and various payday lenders).

Once the data were extracted and sorted from the IRS, we had to find a way to measure a firm's distance from its customers. Two possibilities emerged, compile firm counts in a given ZIP code, or use Geographic Information Software (GIS) to do a more sophisticated analysis. We decided GIS software was most complete, as GIS plots addresses on a map and allows for distances between firm locations to be measured. Due to time constraints, we were unfortunately forced to abandon the use of arcGIS and rely solely on the count of tax preparation firms in a given ZIP. This variable is denoted as RALScout. RALScout explains how many firms in each ZIP code offer RALs, forcing us to making the assumption that people are restricted to the tax services provided in the ZIP code they live in. This assumption loses the more realistic explanatory power that people travel outside their ZIP to consume goods, but is the most practical option second to GIS.

Recommendations for future study in the distribution of firms and their effects on EITC take-up include working through our GIS problems and using firm count in all adjacent ZIP codes identified through GIS buffers. See Appendix 2 for further GIS explanation.

### Taxpayer Characteristics

To determine taxpayer characteristics, the IRS—Stakeholder Partnerships, Education, and Communication (SPEC) Return Information Database was used. This database summarizes original tax return data and provides counts of individual characteristics for all U.S. ZIP codes (Berube 2006). Initiative for using this data was taken from Alan Berube's 2006 literature documenting findings of his EITC analysis study, in which Berube used SPEC data for EITC trends and influence between the years 2000 and 2003. As mentioned in the literature review of Berube's study, the SPEC database groups tax returns into market segments that divide return

data into separate groupings such as EITC received or low-income filers, and then reports tax payer characteristics such as married, or head of household for each grouping. For our purposes market segments used were all individual returns (variables denoted with surname tt), returns for which the taxpayer received EITC (variables denoted with surname et), and returns filed by EITC eligible or potential recipients (variables denoted with surname p).

SPEC data is not freely available to the public, and contact with Berube was made to retrieve SPEC data tables he previously used. Taxpayer characteristics in these tables were available for the years 2000 and 2005, and thus become our only data with any yearly variation. In our first model, SPEC variable total number of returns that received EITC (variable etTOTRET) is used as an explanatory variable to determine firm count in a ZIP code or, a firm's decision to locate. Total number of returns for which EITC was received (variable etTOTRET) also serves as our dependent variable for the second model where explanatory census variables and number of firms in a ZIP code are used to determine EITC-take up. SPEC table data is in count number forms, so generating rates for some variables was necessary and possible by using total individual returns as a divisor.

Further work on determining EITC take-up rates would benefit from time spent cleaning out unnecessary variables from the SPEC data and making these large SPEC files easier to handle. More in-depth analysis would then be possible by taking a closer look at very detailed characteristics listed in the SPEC data such as married EITC filers filing jointly verse single EITC filers. Taking time to further fine tune the data will also make STATA easier to work with, as variables will be easier to identify when running multiple regressions. In addition, finding firm count and census data across different years would allow more use out of this SPEC data, which does have yearly variation and could be used in further time-series analysis.

## Census

Demographic data, such as population, ethnic, education and housing figures, were derived from the U.S. Census Bureau for all available Oregon, Washington and Idaho ZIP codes. Generating “Quick Tables” using Sample Data – Census 2000 Summary Files 3 and 4, provided housing and population values for a portion of the three states’ five digit ZIP codes. Using these values, we were able to calculate explanatory variables, in the form of total counts and rates, for our econometric models (U.S. Census Bureau).

One deficiency in using Census 2000 data was that we were not able to obtain each desirable demographic figure for every five-digit ZIP code of concern. For many of the ZIP codes in question, data was not available through the universal summary files. To adjust for this introduced inaccuracy, we constructed rates of variables along with universal counts for use in the models, and accounted for the increased sample error. Another deficiency is the age of this census data. Almost eight years old, the data does not easily align with the 2005 SPEC data nor the 2007 firm count IRS data.

## **Methodology**

In order to generate regressions which would lead us to the most reliable conclusion concerning our data, we incorporated various functional forms of key variables into a straight forward ordinary least squares format. With each of our regressions, we initially included only one exogenous variable while incrementally adding additional variables. This allowed us to observe the marginal explanatory power of certain variables within the models. We also incorporated models to explain the proportion of EITC received (eitcrate), the amount of Refund Anticipation Loans obtained with the EITC (etral), and the proportion of people who are

potentially eligible for the EITC that collect the EITC (potrate). Using five different dependent variables—ralscount, ettotret, eitcrate, etral and potrate—we constructed the equations to rationalize the type of connection between tax preparation firm location and the rate at which the earned income tax credit is filed for in Idaho, Oregon and Washington. See Appendix 3 for comprehensive STATA .do file.

To explain frequency and location of tax preparation firms within a given ZIP code, we used the variable ralscount—or the number of firms in a ZIP code—as the dependent variable, and total population (totpop), median household income (medianhhinc), the number of persons with incomes below the poverty level (inc99povertylevel), median age (medianage), high school education (highgrad), median age squared (age2) and education squared (educ2) to provide a quadratic element to the model, and finally the number of tax returns receiving the EITC (ettotret) as independent variables. Starting with total population as the only explanatory variable and incrementally adding in median household income and the number of people below the poverty line, we were able to explain 55 percent of the variation in the number of tax preparer firms in a ZIP code. Every coefficient at this point is significant at the 99 percent confidence level, with an increase of one firm being explained by 75,000 more people in a zip code, a \$20,000 decrease in the median household income and an increase of 3,000 people below the poverty level. Controlling for demographic factors, we then added in the total number of tax filers receiving the EITC, increasing explanatory power to 55 percent, and generating the result that 138 more EITC returns leads to an increase of one more tax preparer firm in a ZIP code. Performing F-test of significance of the coefficients in the fourth regression of table 1 yields the result that the coefficients on age squared, median age, education squared and high school

education are not significantly different from zero at the 5 percent significance variable, and therefore do not explain any significant variation in the number of firms within a ZIP code.

Table 1: Firms' Decision to Locate

	(1)	(2)	(3)	(4)	(5)	(6)
	#Firms	#Firms	#Firms	#Firms	#Firms	#Firms
Total Pop	1.077e-4 (25.47)**	1.198e-4 (28.35)**	7.49e-5 (8.14)**	9.07e-5 (3.67)**	2.18e-5 (2.14)**	1.021e-4 (6.84)**
MedianHHinc		-4.26e-5 (9.01)**	-2.05e-5 (3.33)**	-2.43e-5 (3.80)**	-8.17e-6 (-1.38)	2.31e-5 (3.72)**
#<poverty			3.79e-4 (5.47)**	3.301e-4 (4.46)**	1.026e-4 (1.46)**	3.119e-4 (4.30)**
Age^2				-1.216e-3 (1.14)		-1.068e-4 (0.81)
Educ^2				-1.89e-9 (2.26)*		-1.56e-9 (2.29)*
Median Age				8.672e-2 (1.02)		
Highgrad				2.45e-5 (0.55)		
EITC Total					1.381e-3 (9.86)**	
Constant	-0.154 (1.90)	1.443 (7.47)**	0.589 (2.40)*	-0.877 (0.52)	5.33e-2 (0.23)	0.759 (2.20)*
Observations	708	708	708	708	707	708
R-squared	0.48	0.53	0.55	0.56	0.61	0.56
Absolute value of t statistics in parentheses						
* significant at 5%; ** significant at 1%						

To explain the number of tax returns receiving the EITC, we used the variable *ettotret*—or the number of tax filers receiving the EITC in a ZIP code—as the dependent variable, and total population (*totpop*), median household income (*medianhhinc*), the percent of persons with incomes below the poverty level (*povertyline*), the number of people who at least graduated from high school (*highgrad*), the number of people of a minority ethnicity (*nonwht*), the average household size (*avghhsize*), and finally the number of tax preparer firms (*ralscount*) as independent variables. Before adding the number of firms into the model, we were able to explain 84 percent of the variation in EITC take-up with significant coefficients on total population, median household income, percent below the poverty line, number of high school graduates, number of minorities, and average household size. Results show that EITC take-up

increases with higher populations, more minorities, more people in the average household, a lower median household income, a decrease in people below the poverty line and a decrease in high school graduates. Controlling for census variables, we then added in the number of preparer firms, increasing explanatory power to 86 percent, and generating the result that more firms significantly explain an increase in the number of people receiving the EITC in a specified ZIP code.

Table 2: EITC Take-up

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Total EITC	Total EITC	Total EITC	Total EITC	Total EITC	Total EITC	Total EITC
Total Pop	0.052 (45.56)**	0.058 (58.52)**	0.058 (58.37)**	0.091 (20.59)**	0.078 (17.12)**	0.064 (12.92)**	0.056 (10.20)**
Median HH inc		-0.021 (18.51)**	-0.022 (14.08)**	-0.021 (13.33)**	-0.021 (14.22)**	-0.024 (15.74)**	-0.021 (12.28)**
%<poverty line			-4.198 (1.57)	-6.998 (2.70)**	-10.300 (4.05)**	-8.984 (3.62)**	-9.144 (3.41)**
Highgrad				-0.058 (7.56)**	-0.049 (6.59)**	-0.023 (2.73)**	-0.018 (2.01)*
NonWhite					0.041 (7.32)**	0.039 (7.14)**	0.030 (5.18)**
Avg. HH size						233.215 (6.32)**	268.641 (6.77)**
#Firms							93.638 (10.79)**
Constant	56.176 (2.54)*	826.493 (18.21)**	948.309 (10.57)**	923.391 (10.69)**	1,002.675 (11.93)**	476.107 (4.08)**	290.019 (2.29)*
Observations	1414	1414	1414	1414	1414	1414	707
#Zipcodes	707	707	707	707	707	707	707
R^2 Overall	0.73	0.81	0.81	0.82	0.83	0.84	0.86
Absolute value of z statistics in parentheses							
* significant at 5%; ** significant at 1%							

To explain the rate at which tax filers receive the EITC, generated as eitrate, we used the percentage of the population living in urban areas (perurban), the percentage of people who at least graduate high school (perhihgrad), the percentage of minorities (pernonwht), the total number of tax returns filed through a volunteer site (ttvita), and the number of tax preparation firms (ralscount) as regressors. In doing so, we were able to explain 24-25 percent of the variation in the rate of EITC take-up and produce coefficients which were all significant.

Particularly interesting was the finding that a general increase in the number of firms significantly explains an increase in the rate at which the EITC is received by those filing their taxes, and that an overall increase in the number of returns filed through volunteer sites is correlated with a decrease the EITC take-up rate. The number of tax preparer debt inquiries (etral), which describes the number of loans taken out by tax filers from preparation firms in order to receive a portion of their EITC early in exchange for a fee, allowed us to generate a final model to view the unwanted increase in such loans caused by an increase in the number of preparer firms. Generating this model and controlling for demographic variables, we find that it would take an additional 30 firms to significantly explain an increase of one more loan taken out to receive the EITC early, which correlates the location of such firms with the undesired increase in refund anticipation loans.

Table 3: EITC Take-up Rate and Refund Anticipation Loans

	(1)	(2)	(3)	(4)	(5)	(6)
	EITC rate	EITC rate	EITC rate	EITC rate	#Loans	#Loans
Urban%	-0.038	-0.037	-0.045	-0.044		
	(10.65)**	(10.27)**	(7.92)**	(7.79)**		
Hihgrad%	-0.029	-0.31	-0.024	-0.026		
	(4.14)**	(4.37)**	(2.38)*	(2.54)*		
Nonwht%	0.222	0.226	0.208	0.214		
	(16.49)**	(16.70)**	(11.04)**	(11.26)**		
Total Pop					0.021	0.016
					(14.51)**	(8.63)**
Median HH inc.					-0.006	-0.005
					(14.69)**	(8.64)**
%<poverty line					-2.731	-2.184
					(3.78)**	(2.44)*
Highgrad					-0.019	-0.015
					(7.65)**	(5.01)**
Nonwht					0.022	0.014
					(13.75)**	(7.20)**
Avg. HH size					51.760	57.481
					(4.82)**	(4.33)**
#Firms			5.02e-3	5.65e-3		30.173
			(4.53)**	(4.93)**		(10.40)**
Total Volunteer		-3.9e-5		-1.16e-4		
		(2.55)*		(2.10)*		
Constant	0.149	0.150	0.152	0.152	155.390	78.402
	(35.30)**	(35.46)**	(25.18)**	(25.24)**	(4.58)**	(1.85)

Observations	1414	1414	707	707	1414	707
R-squared	0.23	0.24	0.24	0.25	0.71	0.75
Absolute value of t statistics in parentheses						
* significant at 5%; ** significant at 1%						

The proportion of potentially eligible people who actually receive the EITC is explained by the regressions in table 4, which use census, SPEC and firm location variables as regressors; total population, median household income, average household size, percentage of people below the poverty line, high school education, age squared, returns filed by volunteer sites and the number of tax preparation firms. Significant results reveal that non-profit tax preparation activity positively affects the take-up rate of the credit by eligible people, and that the presence of for-profit firms adversely affects this proportion.

Table 4: Potential EITC and Volunteer Impact

	(1)	(2)	(3)	(4)	(5)	(6)
	Potential	Potential	Potential	Potential	Potential	Potential
Total Pop	-1.02e-8 (3.60)**	-1.07e-8 (3.83)**	-1.08e-8 (3.86)**	-8.74e-9 (2.99)**	-9.02e-9 (3.16)**	-9.16e-9 (3.16)**
MedianHHinc	8.48e-9 (9.53)**	9.02e-9 (10.20)**	8.97e-9 (10.07)**	8.06e-9 (8.79)**	8.56e-9 (9.43)**	8.49e-9 (9.31)**
Avg. HH size	-4.16e-5 (16.05)**	-4.25e-4 (16.57)**	-4.23e-4 (16.53)**	-4.17e-4 (16.11)**	-4.26e-4 (16.66)**	-4.24e-4 (16.52)**
%<poverty	6.03e-6 (4.08)**	6.84e-6 (4.66)**	6.63e-6 (4.50)**	5.84e-6 (3.95)**	6.65e-6 (4.54)**	6.44e-6 (4.38)**
Highgrad	1.28e-8 (2.59)*	1.30e-8 (2.67)*	1.33e-8 (2.72)*	1.21e-8 (2.45)**	1.22e-8 (2.52)*	1.25e-8 (2.56)*
Age^2	-1.24e-7 (6.04)**	-1.17e-7 (5.78)**	-1.19e-7 (5.82)**	-1.27e-7 (6.18)**	-1.20e-7 (5.94)**	-1.22e-7 (5.98)**
Volunteer Total		2.57e-7 (4.71)**			2.63e-7 (4.83)**	
Volunteer EITC			2.12e-6 (3.85)**			2.23e-6 (4.05)**
#Firms				-9.24e-6 (1.94)*	-1.03e-5 (2.20)*	-1.09e-5 (2.31)*
Constant	0.002 (16.38)**	0.002 (16.39)**	0.002 (16.34)**	0.002 (16.53)**	0.002 (16.58)**	1.65e-3 (16.55)**
Observations	707	1414	1414	707	1414	1414
R-squared	0.40	0.42	0.41	0.40	0.42	0.42
Absolute value of t statistics in parentheses						
* significant at 5%; ** significant at 1%						

In conclusion, the methodology used to generate the above regressions is able to rationalize that given a tax preparation firm's decision to locate in a given ZIP code, an additional firm leads

to an increase in the total number of people receiving the EITC, the take-up rate of the tax credit, and the number of Refund Anticipation Loans applied for in a given ZIP code.

## Conclusion

Our econometric models and datasets allowed us to examine a wide range of EITC related issues. Building off past research, we expected tax preparation firms to locate in areas where a large percentage of the population qualifies for EITC credit. A major motivation for this behavior is to tap into the Refund Anticipation Loan market. We also anticipated a positive correlation between tax firms and the overall EITC take-up rate. Lastly, we predicted there would be a positive correlation between tax firms and the number of EITC refunds distributed through RALs. The following sections examine each econometric model in detail and outline the implications of our results.

### Model One: Firm Distribution

$$\text{Number of Firms} = (0.053) + (0.0000218)TP + (-0.00000817)MHI + (0.0001026)PL + (0.0013805)EITC + \mu$$

$$R^2 = 0.6059$$

$$N = 707$$

Where:

- TP = total population
- MHI = Median household income
- PL = Number below poverty line
- RAL = Firm count

As we hypothesized, a firm's decision to locate in a given ZIP code is heavily influenced by the number of low-income people and EITC returns in the area. A 1,000 EITC recipient increase translates into 1.3 more firms choosing to locate in the ZIP code. Furthermore, 10,000

more people below the poverty line results in 1.026 more tax firms. Thus, a firm is 10 times more likely to locate in a ZIP code with high EITC returns than it is with only a large amount of the population under the poverty line. This suggests that firms are more interested in locating in areas with established EITC take-up rates, not ZIPs with only a large impoverished population.

#### Model Two: EITC Take-Up

$$\# \text{ of EITC} = (290.01) + (0.056)TP + (-0.0205)MHI + (-9.144)PL + (-0.018)HG + (0.0302)NW + (268.64)AHS + (93.63)RAL + \mu$$

$$R^2 = 0.8637$$

$$N = 707$$

$$\text{Rate of EITC} = (0.152) + (0.208)\%NW + (0.0053)RAL - (0.045)\%U - (0.024)\%HS + \mu$$

$$R^2 = 0.24$$

$$N = 707$$

- TP = total population
- MHI = Median household income
- PL = Percent below poverty line
- HG = High school grad or better
- NW = Number non-white
- AHS = Average household size
- RAL = Firm count
- U = Urban ZIP

Our assumptions regarding increased EITC take up rates in areas with more tax firms was supported by our results. We found that a one RAL offering tax firm increase yields 93.63 more EITC recipients, and a one percent increase in the number of people below the poverty line translates into a 9.14 increase in EITC take-up. Thus, an increase of one tax firm increases the take-up rate 10 times greater than a comparable increase in the number of people below the poverty line. To increase the robustness of our results we ran an additional take-up model with

all variables represented as rates and found that more firms do indeed increase the EITC take-up. This suggests that the profit incentive for firms to seek out customers through advertisements and conspicuous locations is very successful in increasing the number of EITC recipients. To put this in perspective, for the Eugene Zip code of 97401 there were 1686 potential EITC recipients in 2000, of which 1412 received the credit. From our findings, one more tax firm would increase the take-up rate from 83 percent to 90 percent. With an average EITC credit of \$1587 per person, this 7 percent increase translates into \$147, 591 in Federal aid entering the community.

#### Model Three: EITC Refunds Distributed Through RALs

$$\text{RAL} = 78.402 + (-2.184)\text{PL} + (0.016)\text{TP} + (-0.005)\text{MHI} + (0.014)\text{NW} + (-0.015)\text{HG} + (57.481)\text{AHS} + (30.173)\text{RAL} + \mu$$

$$R^2 = 0.75$$

$$N = 707$$

- TP = total population
- MHI = Median household income
- PL = Percent below poverty line
- HG = High school grad or better
- NW = Number non-white
- AHS = Average household size
- RAL = Firm count

The results of this model suggest that more firms and larger family sizes contribute to a person's choice to take a RAL. However, this result is somewhat misleading since an increase in the average family size is also a large increase in total population. A one firm increase yields 30.2 more RALs and an increase of one more person in the average household translates into 57.481 more RALs. This means that even when you hold constant all other demographic characteristics, simply having the option to take out a RAL will increase the number of refund

loans. There are dual forces behind this result; people desire to take loans and more firms make that an option. In addition, firms have an interest in getting their customers to take loans.

#### Model 4: Potential EITC and Volunteer Impact:

We hoped to find how consumers substitute between volunteer organizations and paid providers, however, our regression results were of questionable accuracy. For example, several of our coefficients had the wrong sign and all the coefficients were practically zero. We believe our lack of solid data on volunteer organizations was to blame. Should we have better data, we can analyze whether non-profits are able to attract consumers away from paid preparers.

From these conclusions it is obvious that tax preparation firms are very successful in increasing the EITC take-up rate in the ZIP codes in which they locate. However, the concern is how much of this Federal aid is then channeled into the hands of for-profit institutions in the forms of RALs and preparation fees. It is clear that non-profits have a difficult time competing with for-profit firms who can offer instant funds for hyperbolic discounters in the form of RALs. According to H&R Block, it costs \$165.00 in preparation fees and \$66.04 for the near average \$1500 EITC credit with a RAL, which totals \$231.04 in tax preparation fees. This reduces the average EITC refund of \$1587 to \$1355.96. Thus, 14.56 percent of EITC funds end up in the hands of for-profit businesses.

While nearly 15 percent of the funds from this federal aid program do not end up in hands of the working poor, for-profit firms are motivated to ensure that every eligible EITC recipient receives the funds they are entitled to. On average 222 persons per ZIP receive their EITC from for-profit tax preparation firms, whereas only 12 people gain an EITC from a volunteer site. The

net social benefit from the quantity of customers receiving the EITC far outweighs the 15 percent of aid money that leaves the system. However, it is important to note that the EITC in its current form is not a free system. There are many inherent costs attributed to increasing the take-up rate, which mean fewer dollars end up in the hands of the poor. Furthermore, it would be nearly impossible to shut these paid preparers out of the market at this time. These firms are making millions of dollars from preparation fees and loans. Thus, they have both the money and motivation to ensure their continued access to EITC customers.

## **Recommendations**

While we were able to identify nearly 90 percent of the characteristics that explain EITC take-up rate, additional data and regression analysis could raise that number even higher. Our biggest difficulty in explanatory power was the inability to find time-series data, particularly with respect to firm counts per ZIP code. The data obtained through the IRS, which supplied us with the addresses of each tax firm, only kept current information, meaning it was unknown when a particular tax preparer opened for business. Time-series variation would allow us to determine the impact an additional firm has on the EITC. Furthermore, due to the low population density in the three states that were chosen, many ZIP codes had no census information. This forced us to cut a significant number of census observations.

The SPEC data P-group indicating potential EITC recipients or eligible EITC recipients was created by Berube for his own work and it is unclear exactly how he calculated this potentiality. In future research, this process needs to be determined, or else an independent means of identifying EITC eligibility may be created and used as a measure for potential EITC recipients. In addition, the SPEC data contained nearly 150 variables which we were unable to

fully explore due to time constraints. This data could prove especially useful in determining the EITC take-up among very specific groups, such as single parents, those with large families and the elderly. Since the EITC is a national program, it would also be beneficial to expand our scope to include the entire United States. There are unique characteristics to the Northwest, such as a small minority population and a relatively low per capita income, which make the area not representative of the entire country.

As indicated in our regression results, 20 percent of those who file through a volunteer site receive the EITC compared to only 14 percent who file with a paid preparer. These statistics reflect the lower median-income of individuals who have their taxes prepared by volunteer sites. This means non-profits are able to provide service to a very niche market that is the most in-need: those who cannot afford to pay for tax preparation fees. However, if these volunteer sites want to expand their client base, they must be able to offer a broader range of services. Theory suggests that one reason hyperbolic discounters choose for-profit preparation firms is the firm's ability to offer RALs. The fact that a consumer can walk into a tax firm and walk out with their EITC refund the same day—regardless of the expense—is the key reason they choose for-profit firms. If this is true, then no matter how many free volunteer sites are accessible, nor how well they advertise their services, consumers will still pick a business that offers RALs. While this makes the goals of non-profit organizations such as United Way more difficult to obtain, by no means does it make them impossible. It must be realized that we cannot control the behavior of taxpayers, but that there may be possible ways to influence their actions or meet their demands. There are possible avenues by which an organization such as United Way of Lane County could affect the choices made by low to moderate income workers through educational programs and

public outreach, yet if offering RALs increases EITC take-up rates, then non-profit organizations should attempt to devise a method to provide such immediate funds.

A possible policy suggestion in pursuing the ability to offer services like Refund Anticipation Loans, would be to lobby the support of banks, credit unions and other financial institutions. If these organizations underwrote RALs they would be doing a service to the community and could derive a profit. Currently, the majority of firms offer RALs at a 36 percent APR, which is the maximum that can be charged in the state of Oregon. This leaves financial institutions with plenty of room to both lower rates and ensure they derive a return on their investment. Furthermore, it is reasonable to assume that some consumers who desire a RAL would be willing to accept a loan for a portion of their entire EITC refund, which would lower the monetary outlay required by the institutions that underwrite the loans. In addition, many of these hyperbolic discounters are also unbanked, so this interaction with a financial institution would provide an opportunity for LMI individuals to begin relationships with a bank and establish checking and savings accounts for the first time.

Targeting and reaching people who are potentially eligible to receive the EITC decreases the distance (in time and space) between them and volunteer tax sites, and increases the probability that they will participate in such programs. Possible locations that United Way of Lane County could target with educational outreach programs include Birth to Three and local high schools. Birth to Three, and related community human services groups, attracts a population commonly comprised of single, low-income parents, and in many high schools, there exists the possibility of reaching seniors (18 years old) who are facing immediate uncertainties of their future financial support, incomes, career and family status. Such settings offer great opportunity

for organizations like United Way of Lane County to educate the public of their services and provide incentives to utilize them.

## References

- Berube, A. (2006) "The New Safety Net: How the Tax Code Helped Low-Income Working Families During the Early 2000s." Metropolitan Policy Program, The Brookings Institution. Feb. 2006.
- Berube, A. and Forman, B. (2001) "A Local Ladder for the Working Poor: The Impact of the Earned Income Tax Credit in U.S. Metropolitan Areas." The Brookings Institute. Sept. 2001.
- Berube, Kim, Benjamin and Burns. (2002) "The Price of Paying Taxes: How Tax Preparation and Refund Loan Fees Erode the Benefits of the EITC." The Brookings Institute. May 2000.
- Caskey, J. (1994) "Who Has a Bank Account and Who Doesn't: 1977 and 1989." *Eastern Economic Journal*. 20 (1). Winter 2004.
- Greene, Rhine and Toussaint-Comeau. Working Paper, (2003) "The Importance of Check-Cashing Businesses to the Unbanked: Racial/Ethnic Differences." Federal Reserve Bank of Chicago. August 2003.
- Greenstein, Robert. (2003) "The New Procedures for the Earned Income Tax Credit," *Center on Budget and Policy Priorities* 2003. 1-19.
- Holtzblatt, McCubbin and Gillette. (1994) "Promoting Work through the EITC." *National Tax Journal*. 47 (3): 591-607. Sept. 1994.
- Kopczuk, Wojciech, Pop-Eleches, Cristian. (2007) "Electronic Filing, Tax Preparers and Participation in the Earned Income Tax Credit." *Journal of Public Economics* 91(2007): 1351-1367.
- Neumark, D. and Wascher, W. Working Paper, (2000) "Using the EITC to Help Poor Families: New Evidence and a Comparison with the Minimum Wage." National Bureau of Economic Research. March 2000.
- Scholz, John K. (1994) "The Earned Income Tax Credit: Participation, Compliance, and Antipoverty Effectiveness," *National Tax Journal* 47(1): 63-87.
- US Census Bureau. "Data Sets: American FactFinder." <<http://factfinder.census.gov/servlet/DatasetMainPageServlet>>.
- Vermilyea, T. & Wilcox, J. Working Paper, (2002) "Who is Unbanked and Why: Results from a Large, New Survey of Low- and Moderate-Income Adults." Federal Reserve Bank of Chicago. May 2002.
- Welfare Peer Technical Assistance Network. (2006) "The Earned Income Tax Credit: Supporting the Working Poor in Creating Long-term Self-sufficiency." U.S. Department of Health and Human Services. Summer 2006.

## Appendix 1

### SPEC tax-payer characteristics and Census data summary statistics

Variable Description	Variable	# of Obs	Mean	Std. Dev.	Min	Max
Total Tax Returns	tttotret	3085	3180.331	5024.808	0	27850
EITC Returns	ettotret	3085	401.3981	659.4816	0	4417
Average EITC Sum In Dollars	avgeitc	2654	1586.708	293.828	272.125	2832.25
Total Returns With Child Tax Credit	ttctc	3085	621.2674	1023.662	0	7218
EITC Returns With Child Tax Credit	etctc	3085	126.6373	218.726	0	1761
Total Tax Returns Filed By Paper	ttpaper	3085	1813.021	2986.723	0	19749
EITC Tax Returns Filed By Paper	etpaper	3085	135.5002	223.0233	0	1644
Total Tax Returns E-Filed	ttelf	3085	1367.042	2309.351	0	15289
EITC Tax Returns E-Filed	etelf	3085	264.7958	467.9942	0	3583
Total Returns Filed By Self	ttself	3085	1568.949	2622.515	0	16784
EITC Returns Filed By Self	etself	3085	163.5015	266.8864	0	2021
Total Returns Filed By Paid Provider	ttpaid	3085	1545.547	2352.412	0	13415
EITC Returns Filed by Paid Provider	etpaid	3085	221.9971	377.8975	0	2960
Total Returns Filed By Volunteer	ttvita	3085	10.0175	66.75287	0	2340
EITC Returns Filed By Volunteer	etvita	3085	1.388655	8.493627	0	208
Total Preparer Debt Inquiries	ttal	3085	172.3598	325.6147	0	2469
EITC Preparer Debt Inquires	erral	3085	90.53971	176.9159	0	1558
Total Returns with EITC Eligibility	ptotret	1536	402.9674	668.6659	0	4357
Total Population	totpop	1415	14208.95	12985.4	58	64181
Percent Below Poverty Line	povertyline	1415	11.95445	6.627361	0	69.32409
Number Below Poverty Line	inc99pover~l	1415	1524.378	1658.527	0	9442
Number Of Non-White People	nonwht	1415	1744.72	2671.829	0	26452
Number of People With H.S. Diploma (Or Better)	highgrad	1415	8329.768	7742.962	42	35289
Per Capita Income	percapinc	1415	20330.68	6524.864	1531	53799
Median Household Income	medianhhinc	1415	41511.5	11607.96	10465	96028
Number of Urban People	urban	1415	11681.88	13089.01	0	64181
Number of Firms That Offer RALs	ralscout	3088	0.64022	1.531368	0	12

## **Appendix 2**

### GIS explanation

The first step in using GIS is to convert a standard mailing address into a latitude-by-longitude address, which is known as “geo-coding”. We geo-coded all firm addresses obtained through the IRS website, and, once geo-coded, latitude-by-longitude coordinates were imported into arcGIS, a program that plots coordinates onto a map in their geographic locations. In short, a map is produced, covered by dots that represent the location of tax preparation firms. The benefit of using GIS software is it allows for a particular firm’s distance from other ZIP codes or cities to be calculated, a function not available if we only used a firm count for each ZIP code.

The economic theory behind using arcGIS to plot firm locations is that people do not only consume goods in the ZIP code they live in, but travel across ZIP lines for daily tasks such as shopping and work. This theory also applies to tax preparation. Individuals may choose to have their taxes prepared in adjacent ZIP codes from their home addresses, thus GIS allows identification of a firm’s proximity to a taxpayer, regardless of ZIP boundaries. From this theoretical base, we decided to employ a GIS command that counts not just the firms in one ZIP, but instead total firms in a five or 10 mile radius of a taxpayer’s residence. This function accounts for the potential misrepresentation of certain ZIP codes that have no tax-firms, such as rural areas, which would produce a firm count of zero, when in actuality they are multiple firms within five miles of that ZIP code. GIS smoothes the firm count in all ZIP codes, and more accurately represents a taxpayer’s options for tax preparation locations.

To determine firm count using GIS, we used a mathematical command that finds the center of each ZIP—known as the centroid—and from there made the assumption that the entire population of a given ZIP lives in the centroid. While this is not a completely accurate assumption, it is the most practical method for counting taxpayers. Once the centroid was

established, it provided us with the necessary starting point to begin measuring distances between residences and tax preparation firms. We then created several radii—known as buffers—and recorded the number of firms within each buffer from the centroid. The buffers we used were one, three, five, ten and thirty mile distances from the centroid. The 10 and 30 mile buffers were chosen to capture the distance that many people in rural areas would be forced to travel for tax preparation services.

Unfortunately, it proved impossible to identify which firms belonged to which ZIP centroids. The software did not allow buffers to be placed around single ZIP codes, and would not uniquely identify which ZIP codes the firms belonged to. In other words, we were only able to ascertain the total number of firms that were captured at a given distance from the centroid of all ZIP codes, and not the specific ZIP code a firm is associated with.

### Appendix 3

STATA .do file

```
clear
set mem 4000k
use "E:\STATA\specfreqcen.dta"
```

```
replace ralscount=0 if ralscount==.
replace ralscount=. if year==2000
```

```
gen nonwht = blkfam + aminalsknat + asian + hapacis + other + races
gen highgrad = m12thgradenodiploma + mhighschoolgraduateincludesequiv +
msomecollegelessthan1year + msomecollege1ormoreyearsnodegree + massociatedegree +
mbachelorsdegree + mmastersdegree + mprofessionalschooldegree + mdoctoreddegree +
f12thgradenodiploma + fhighschoolgraduateincludesequiv + fsomecollegelessthan1year +
fsomecollege1ormoreyearsnodegree + fassociatedegree + fbachelorsdegree + fmastersdegree +
fprofessionalschooldegree + fdoctoreddegree
```

```
drop blkfam aminalsknat asian hapacis other races m12thgradenodiploma
mhighschoolgraduateincludesequiv msomecollegelessthan1year
msomecollege1ormoreyearsnodegree massociatedegree mbachelorsdegree mmastersdegree
mprofessionalschooldegree mdoctoreddegree f12thgradenodiploma
fhighschoolgraduateincludesequiv fsomecollegelessthan1year
fsomecollege1ormoreyearsnodegree fassociatedegree fbachelorsdegree fmastersdegree
fprofessionalschooldegree fdoctoreddegree
```

```
gen targetage = m18and19years + m20years + m21years + m22to24years + m25to29years +
m30to34years + m35to39years + m40to44years + m45to49years + m50to54years + m55to59years +
m60and61years + m62to64years + f18and19years + f20years + f21years + f22to24years +
f25to29years + f30to34years + f35to39years + f40to44years + f45to49years + f50to54years +
f55to59years + f60and61years + f62to64years
gen age2 = medianage^2
gen educ2 = highgrad^2
gen etvitarate = etvita/ettotret
```

```
drop f85yearsandover f80to84years f75to79years f70to74years f67to69years f65and66years
m85yearsandover m80to84years m75to79years m70to74years m67to69years m65and66years
funder5years f5to9years m5to9years munder5years
drop _merge
drop f11thgrade f10thgrade f9thgrade f7thand8thgrade f5thand6thgrade fnurseryto4thgrade
fnoschoolingcompleted m11thgrade m10thgrade m9thgrade m7thand8thgrade m5thand6thgrade
mnurseryto4thgrade mnoschoolingcompleted
drop f15to17years f10to14years m15to17years m10to14years
drop p2agi_1x p1agi_1x p0agi_1x p2agi_75 p1agi_75 p0agi_75 p2agi_60 p1agi_60 p0agi_60
p2agi_50 p1agi_50 p0agi_50 p2agi_40 p1agi_40 p0agi_40 p2agi_35 p1agi_35 p0agi_35 p2agi_30
p1agi_30 p0agi_30 p2agi_25 p1agi_25 p0agi_25 p2agi_20 p1agi_20 p0agi_20 p2agi_15 p1agi_15
p0agi_15 p2agi_10 p1agi_10 p0agi_10 p2agi_5 p1agi_5 p0agi_5 p2agi_0 p1agi_0 p0agi_0 p2etop
p1etop p0etop p2ptoe p1ptoe p0ptoe p2rep_p p1rep_p p0rep_p p2rep_e p1rep_e p0rep_e p2elf
p1elf p0elf p2paper p1paper p0paper p2oth_e p1oth_e p0oth_e p2oth_p p1oth_p p0oth_p p2se_el
p1se_el p0se_el p2se_pa p1se_pa p0se_pa p2pr_el p1pr_el p0pr_el p2pr_pa p1pr_pa p0pr_pa
p2irspr p1irspr p0irspr p2tce p1tce p0tce p2vita p1vita p0vita p2totvol p1totvol p0totvol p2paid
p1paid p0paid p2self p1self p0self p2fsc3 p1fsc3 p0fsc3 p2fsc2 p1fsc2
```

```
drop p0fsc2 p2fsc1 p1fsc1 p0fsc1 p2ral p1ral p0ral p2esttx p1esttx p0esttx p2dirdep p1dirdep
p0dirdep p2sbal p1sbal p0sbal p2cbal p1cbal p0cbal p2sref p1sref p0sref p2cref p1cref p0cref
p2eicctc p1eicctc p0eicctc p2cctc p1cctc p0cctc p2ceicc p1ceicc p0ceicc p2new p1new p0new
```

```
gen eitcrate = ettotret/tttotret
gen avgeitc = etseicc/ettotret
```

```
reg ralscount totpop medianhhinc medianage avghhsize nonwht
gen ralscl=log(ralscount+1)
gen totpopl=log(totpop)
gen nonwhtl=log(nonwht)
gen highgradl=log(highgrad)
gen medianagel=log(medianage)
gen inc99povl=log(inc99povertylevel)
gen avghhsizel=log(avghhsize)
gen medagesq=medianage^2
gen perurban = urban/totpop
gen perhighgrad = highgrad/totpop
gen perralsc = ralscount/totpop
gen pernonwht = nonwht/totpop
```

```
gen totpot = p0seicc + p1seicc + p2seicc
gen potrate = ettotret/totpot
```

```
reg ralscount totpop
reg ralscount totpop medianhhinc
reg ralscount totpop medianhhinc inc99povertylevel
reg ralscount totpop medianhhinc inc99povertylevel age2 educ2 medianage highgrad
test totpop=medianhhinc=inc99povertylevel=age2=educ2=medianage=highgrad=0
test age2
test educ2
test medianage
test highgrad
test totpop
test medianhhinc
test inc99<povertylevel
reg ralscount totpop medianhhinc inc99povertylevel ettotret
```

```
xtreg ettotret totpop, re i(zipcode)
xtreg ettotret totpop medianhhinc, re i(zipcode)
xtreg ettotret totpop medianhhinc povertyline, re i(zipcode)
xtreg ettotret totpop medianhhinc povertyline highgrad, re i(zipcode)
xtreg ettotret totpop medianhhinc povertyline highgrad nonwht, re i(zipcode)
xtreg ettotret totpop medianhhinc povertyline highgrad nonwht avghhsize, re i(zipcode)
reg ettotret totpop medianhhinc povertyline highgrad nonwht avghhsize ralscount
```

```
reg eitcrate perurban perhighgrad pernonwht
reg eitcrate perurban perhighgrad pernonwht ttvita
reg eitcrate perurban perhighgrad pernonwht ralscount
reg eitcrate perurban perhighgrad pernonwht ralscount ttvita
```

```
reg etral totpop medianhhinc povertyline highgrad nonwht avghhsize
reg etral totpop medianhhinc povertyline highgrad nonwht avghhsize ralscount
```

```
sort zipcode year
reaplace ralscount = ralscount[_n+1] if year==2000

reg potrate totpop medianhhinc avghhsize povertyline highgrad age2
reg potrate totpop medianhhinc avghhsize povertyline highgrad age2 ttvita
reg potrate totpop medianhhinc avghhsize povertyline highgrad age2 etvita
reg potrate totpop medianhhinc avghhsize povertyline highgrad age2 ralscount
reg potrate totpop medianhhinc avghhsize povertyline highgrad age2 ralscount ttvita
reg potrate totpop medianhhinc avghhsize povertyline highgrad age2 ralscount etvita
```