The Outbreak of a Tax Break: Essays on the Participation and Impact of the Saver's Credit Across Time and Distance

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Victoria L. Bryant
Master of Arts
George Mason University, 2008
Bachelor of Arts
Virginia Polytechnic Institute, 2003

Director: Sita N. Slavov, Professor of Public Policy
Schar School of Policy and Government

Spring Semester 2020
George Mason University
Fairfax, VA

ACKNOWLEDGEMENTS

I would like to thank the many friends, relatives, and supporters who have made this happen. I would also like to thank Sita Slavov, Anh Pham, Siona Listokin, Shanti Ramnath, Kevin Pierce, Barry Johnson, Gordon Akudibillah, Christopher Bateman, and Julie Bryant for assistance, review, helpful comments, and general support. Finally, thanks go out to the Fenwick Library for providing a clean, quiet, and well-equipped repository in which to work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

THE OUTBREAK OF A TAX BREAK: ESSAYS ON THE PARTICIPATION AND IMPACT OF THE SAVER'S CREDIT ACROSS TIME AND DISTANCE

Victoria L. Bryant, Ph.D.

George Mason University, 2020

Dissertation Director: Dr. Sita N. Slavov

Shortly after the introduction of the Saver's Credit in 2002, the research world was quick to explore its impacts, all imposing the inherent assumption that taxpayers' awareness, and subsequent utilization, would be relatively swift and void of any spatial dimension. However, an ever-growing body of work challenges these assumptions finding that communities play a significant role in the dissemination of and response to information. Now that nearly two decades have passed, I expose the locational heterogeneity, spatial dynamics, and long run impact of the Saver's Credit. Through thematic mapping, spatial autocorrelation, and spatial regression modeling using the population of tax returns from 1999 to 2013, I uncover the influence that neighborhood characteristics have on participation. I find participation is not just spatially autocorrelated but strongly associated with the concentration of married households and the use of a paid preparer. I then measure the impact of the credit on retirement contributions through a Regression Discontinuity model to show that overall contributions have improved, between 12 to 24 percent per person per year, a result that is relatively sensitive to economic cycles.

CHAPTER 1 - INTRODUCTION

The Economic Growth and Tax Relief Recovery Act of 2001 (EGTRRA) is touted as one of the largest tax bills in the past thirty years and arguably ranked among the highest advocating for expansion of retirement savings. Among others, it contained over 40 provisions encouraging participation in retirement savings programs. While there is a significant amount of research to suggest that EGTRRA did improve participation and contributions, there is still more research reporting a continued insufficient level of savings. In a recent report from Gallup's annual Economy and Personal Finance Survey (Saad, 2015) approximately 60 percent of Americans are "moderately" to "very" concerned about not having enough money for retirement, ranking it the top financial concern reported. The Government Accountability Office (GAO) recently reported that just over half of current households aged 55 and older have no retirement savings beyond a Defined Benefits plan through their employer (Jeszeck, 2015). The consequence of insufficient retirement income goes beyond concerns on household expenditures and impacts the level of strain such insufficiency has on social safety net programs.

Among the numerous provisions encouraging participation in retirement savings programs, EGTRRA introduced the Retirement Savings Contributions Credit, commonly referred to as the Saver's Credit. This is a nonrefundable credit to low- and middle-income households who choose to contribute to any qualified retirement savings plan.

The introduction of this credit marked an unprecedented moment in federal legislation. This was the first legislated tax provision to encourage retirement contributions targeted specifically at low- and middle-income households. While the initial design of the credit called for refundability, due to revenue target limitations, the program was converted to be nonrefundable (Gale, Iwry, Orszag, 2004). Although it was originally set to expire in 2005, the Pension Protection Act of 2006, PPA, made this credit permanent and indexed income levels to inflation at increments of $500 (Congressional Research Service, 2006).

By design, the Saver's Credit provides a progressively smaller percentage match for lower and middle-income households to participate in retirement savings programs in order to lower tax liability. Specifically, the credit provides a 50, 20, or 10 percent credit rate on eligible retirement savings contributions of up to $2,000 based on filing status and Adjusted Gross Income, AGI, that can be used to offset any accrued tax liability. Table 1.1 presents these various income notches across several years. Of note, there is a broad band of income that is eligible to receive a 50 percent credit rate, a very narrow band receiving the 20 percent rate, and a somewhat larger band receiving the 10 percent rate. However, because the credit is nonrefundable the theoretical advantage is diminished. Further, given the strict income notches, the plan may introduce unintended consequences of income shifting to qualify.

Table 1.1: Saver's Credit Adjusted Gross Income Thresholds by Marital Status, 2002-2013

Marital Status	Adjusted Gross Income Range per Credit Rate ($) 50%	20%	10%
Married Filing Jointly			
2002-2006	0 - 30,000	30,001 - 32,500	32,501 - 50,000
2007	0 - 31,000	31,001 - 34,000	34,001 - 52,000
2008	0 - 32,000	32,001 - 34,000	34,001 - 53,000
2009	0 - 33,000	33,001 - 34,500	34,501 - 55,500
2010	0 - 33,500	33,501 - 36,000	36,001 - 55,500
2011	0 - 34,000	34,001 - 36,000	36,001 - 56,500
2012	0 - 34,500	34,501 - 36,500	36,501 - 57,500
2013	0 - 36,000	35,501 - 37,500	37,501 - 59,000
Head of Household [1]			
2002-2006	0 - 22,500	22,501 - 24,375	24,376 - 37,500
2007	0 - 23,250	23,251 - 25,500	25,501 - 39,000
2008	0 - 24,000	24,001 - 25,875	25,876 - 39,750
2009	0 - 24,750	24,751 - 27,000	27,001 - 41,625
2010	0 - 25,125	25,126 - 27,000	27,001 - 41,625
2011	0 - 25,500	25,501 - 27,375	27,376 - 42,375
2012	0 - 25,875	25,876 - 28,125	28,126 - 43,125
2013	0 - 26,625	26,626 - 28,875	28,876 - 44,250
Other [2]			
2002-2006	0 - 15,000	15,001 - 16,250	16,251 - 25,000
2007	0 - 15,500	15,501 - 17,000	17 001 - 26,000
2008	0 - 16,000	16,001 - 17,250	17,251 - 26,500
2009	0 - 16,500	16,501 - 18,000	18,001 - 27,750
2010	0 - 16,750	16,751 - 18,000	18,001 - 27,750
2011	0 - 17,000	17,001 - 18,250	18,251 - 28,250
2012	0 - 17,250	17,251 - 18,750	18,751 - 28,750
2013	0 - 17,750	17,751 - 19,250	19,251 - 29,500

[1] Head of Household cutoffs are equal to three-fourths those of married filing jointly
[2] Other includes single, married filing separately, and widowed. Other is equal to one-half those of married filing jointly
Note: The credit is applied at the individual level, not the tax return level.
Source: Internal Revenue Service

Since its introduction, the number of returns claiming the Saver's Credit has hovered around 4 percent of all returns nationally, showing only marginal increases in the last few years, Table 1.2. Those returns that have both income below the cutoff and reported tax liability, i.e. the targeted population from which the credit is attempting to induce more savings, shows participation has grown from 12.04 percent in 2002 to 17.64 percent in 2013. This represents a growth of just under 50 percent. If we redefine eligibility to be those returns with income below the cutoff, having a tax liability, and making retirement contributions, participation rates jump to 57.63 percent in 2004 but only grows 17 percent to end at 67.22 percent in 2013. These rates call into question the credit's ability to induce the desired behavioral change and imply that taxpayers are not just under-utilizing the credit but may not even be aware of its existence.

Putting this into perspective, we can compare these rates to that of the Earned Income Tax Credit (EITC) which is perhaps the most comparable tax incentive program in that it too targets lower income households. The EITC is a means-tested refundable tax credit providing incentives for lower income households to participate in the labor force. This program was established in 1975 with three large expansions in 1986, 1990, and 1993. Estimates put take-up for eligible taxpayers at approximately 70 percent in 1984, 80 to 86 percent in 1990 (Scholz 1994) and 73 to 77 percent in 2005 (Plueger 2009) though exact participation rates are difficult to estimate given the lack of comprehensive observable information on either collected survey or administrative data [1,2].

---

[1] See Currie (2006) and Bhargava and Monoli (2015) for a full review of take up activity.
[2] Determining EITC eligibility on survey data is relatively straightforward but information on receipt is difficult to obtain. Conversely, determining EITC receipt on administrative data is straightforward but information on eligibility is the limitation.

Table 1.2: Number of Tax Returns and Percent Claiming the Saver's Credit, 2002-2013

	2002		2003		2004		2005	
Category	Total Number of Taxpayers	Percent Claiming Saver's Credit	Total Number of Taxpayers	Percent Claiming Saver's Credit	Total Number of Taxpayers	Percent Claiming Saver's Credit	Total Number of Taxpayers	Percent Claiming Saver's Credit
Total	**129,431,998**	**4.10%**	**130,121,189**	**4.07%**	**132,047,569**	**4.01%**	**134,306,525**	**3.94%**
Eligible for credit [1]	45,263,143	11.73%	43,989,887	12.04%	42,507,310	12.44%	41,819,525	12.66%
… And making a contribution [2]	n.a.	n.a.	n.a.	n.a.	9,195,374	57.63%	8,981,590	59.05%
	2006		**2007**		**2008**		**2009**	
Category	Total Number of Taxpayers	Percent Claiming Saver's Credit	Total Number of Taxpayers	Percent Claiming Saver's Credit	Total Number of Taxpayers	Percent Claiming Saver's Credit	Total Number of Taxpayers	Percent Claiming Saver's Credit
Total	**138,339,323**	**3.75%**	**153,544,687**	**3.82%**	**142,445,613**	**4.18%**	**140,490,133**	**4.45%**
Eligible for credit	40,269,872	12.89%	42,249,797	13.88%	40,118,722	14.86%	40,363,436	15.49%
… And making a qualified contribution	8,607,730	60.31%	9,500,032	61.53%	9,630,388	62.19%	10,188,393	61.51%
	2010		**2011**		**2012**		**2013**	
Category	Total Number of Taxpayers	Percent Claiming Saver's Credit	Total Number of Taxpayers	Percent Claiming Saver's Credit	Total Number of Taxpayers	Percent Claiming Saver's Credit	Total Number of Taxpayers	Percent Claiming Saver's Credit
Total	**142,877,036**	**4.29%**	**145,359,282**	**4.40%**	**144,924,478**	**4.78%**	**147,351,299**	**5.03%**
Eligible for credit	41,158,824	14.89%	41,241,890	15.51%	41,551,444	16.67%	42,009,827	17.64%
… And making a qualified contribution	9,588,827	64.19%	9,921,528	64.18%	10,364,686	66.62%	10,844,795	67.22%

n.a. -- Not available

[1] Eligibiligy is defined as having income below the Saver's Credit income cut-off and having a tax liability

[2] These include tax returns that are both eligible and making a contribution into a qualified retirement plan.

Source: IRS Complete Report 2002-2013

Comparatively, participation in the Saver's Credit is significantly lagging. Contributing to the disparity, the EITC is refundable, has been available for 25 years longer, and the IRS actively promotes participation by sending automated notifications to taxpayers who appear to qualify but do not claim EITC benefits.

While Saver's Credit participation shows low uptake on aggregate, such statistics hide the underlying variability and spatial influences occurring across the filing population and speak nothing to the influence the credit has had on improving retirement contributions. Behind these national statistics are a vast array of local distributional responses that indicate a slightly more positive reception.

This research, unlike research that has come before, takes a holistic approach to exploring how the Saver's Credit has changed the landscape of participation and taxpayer behavior. I first exhaustively examine the topography of participation focusing primarily on the Southeast region. I chose the Southeast because it included a mix of urban, suburban, and rural population clusters as well as encompassing much of the east coast to which more than a quarter of U.S. population resides.[3] Then I scrutinize whether the credit has had an impact on taxpayer filing behavior. More precisely, I explore whether the Saver's Credit has encouraged retirement contributions or merely income shifting to qualify. For this, I employ the full population of federal administrative tax return filings matched to information returns reporting retirement contribution behavior for years 1999 through 2013. I break the tax return down to the individual level and match to both Individual Retirement Account (IRA) contributions and Deferred Compensation (DC) information, marking this as one of the most exhaustive datasets developed to explore the influence and impact of the Saver's Credit. The results of my research provide the most comprehensive examination of the Saver's Credit to date and serves as a model for future investigations into tax credit utilization research.

I have divided this research into a four-part analysis. The first three parts, target the evolution of participation to include spatial variation, correlation, and spatial spillovers across the United States, focusing primarily on the Southeast region, to glean additional insights into the concentration and dispersion of participation and explore the correlation between credit participation and retirement savings contribution participation.

[3] In future iterations of this analysis, other regions will be analyzed.

I then measure the spatial spillovers across 5-digit ZIP Codes to explicitly quantify the influence that neighboring regions have on one another. This marks the first study to examine the spatial and serial responses to the Saver's Credit. By relying on the universe of tax administration data aggregated to the 5-digit ZIP Code level and spanning 15 years, from before the credit was introduced to over a decade later, I achieve a detailed picture of not just the dispersion of participation, but also a quantifiable measure of the exact spatial relationship and its impact. I find some areas have shown quite significant improvements in participation both along time and space dimensions while other areas show no discernable improvements along either dimension. I also find that these observed pockets of participation growth are not independent but rather significantly spatially autocorrelated with spillovers playing a significant role in explaining credit participation.

More specifically, Part 1 of this 4-part study, found in Chapter 4, presents the variation and growth in participation, pinpointing locational clusters of (in)activity, and maps the pattern of dispersion over time. It also maps participation in qualified retirement programs over the same horizons and overlays the two maps. From this, I uncover a tight spatial relationship between credit participation and participation in retirement programs through contributions that improves over time and across populations of eligibility. I first locate the initial clustering of Saver's Credit activity, generally concentrated on the borders between Virginia, West Virginia, and North Carolina. Then, I present the year-over-year expansion of participation from these initially isolated clusters to most of the Southeast region. I conduct this same exercise for participation in qualified retirement savings programs and overlay the two to show how they move in tandem over time. Such an endeavor serves as a critical way of discovering areas of initial and continued success of this tax expenditure program, locating areas for potential outreach, and better understanding the geographic direction of credit participation growth.

Part 2 presented in Chapter 5, explicitly measures the spatial distribution of the Saver's Credit participation across this same region to highlight the correlation between localities. Where Chapter 4 draws insight from a thematic mapping of participation and strongly suggests the existence of spatial autocorrelation, Chapter 5 statistically measures that correlation. Here, I primarily rely on the Moran's I statistic to identify and quantify spatial autocorrelation between localities with respect to both Saver's Credit participation and participation through contributions to a qualified retirement plan. Globally, I find a relatively high initial correlation that wanes as time passes, signifying that locational dependence for credit take up is relaxing. I then expound on this by employing the Local Moran's I statistic to identify the specific locations that exhibit higher and lower activity clustering. This offers the added benefit of highlighting the varying distribution of autocorrelation across the region that the Moran's I statistic obfuscates. In so doing, I uncover initially large clusters of highly concentrated behavior focused along the Appalachian region and low participation clusters surrounding urban areas, western Mississippi, and southern Georgia. These acute spatial clusters dissipate over time suggesting that knowledge, and subsequent take-up, of the Saver's Credit is spreading and points to the growing relevance of this credit at which earlier research dismissed.

In Part 3, presented in Chapter 6, I explicitly measure the magnitude and direction of the effect that location has on participation by applying spatial regression models. From the conclusions of Chapters 4 and 5, there is clearly spatial dependence underpinning Saver's Credit participation. In this third study, I compare four different models on the percent of credit participation for each ZIP Code that progressively relaxes the assumption of independence. These models include Ordinary Least Squares (OLS), Spatial Lag model (SAR), Spatial Error Model (SEM), and Spatial Durbin Model (SDM)[4]. In more traditional regression modeling, to include OLS, spatial dependencies are assumed to be zero and thus entirely ignored. Although perfectly reasonable when working with smaller datasets that are randomly sampled across geography, given my dataset's highly geographic granularity this assumption of independence is violated and should not be applied, thus calling into question results of an OLS model. From this research, I find spatial spillovers to be a significant contributor to credit participation. While exact magnitudes dissipate over time, they consistently remain the leading force of observed activity.

In Part 4, presented in Chapter 7, I turn attention to the impact the Saver's Credit on taxpayer filing behavior and retirement contributions. Where before I show a clear linked between credit participation and retirement contributions, this study examines the direction of the relationship. Here, I capitalize on the Saver's Credit sharp income cutoffs to employ a Regression Discontinuity (RD) design to elicit the causal effect of the credit

[4] Throughout Chapter 6 I use language like 'influence', 'effect', and 'drivers' to describe the relationship between participation and various other filing behaviors, to be clear I am explicitly measuring correlation and causation should not be inferred.

on overall retirement contributions. Through this investigation, I show that the expected direction of credit driving contributions is in many regards violated, with an estimated average of 33,524 taxpayers shifting income to qualify for, or for more, the credit, indicating that the identified unintended consequence of income shifting is in fact occurring. I also uncover suggestive evidence that the act of indexing income cutoffs to inflation may have the opposite effect of making it harder for taxpayers to manipulate income to perfectly maximize the credit benefit. After controlling for this bunching behavior, I produce an estimated Saver's Credit impact ranging from an average $55 to $84 of more retirement contributions in response to the credit.

Beyond contributing to the field of study exploring the causal effect of the Saver's Credit, discussed in detail in Chapter 2, this research also contributes to the field of tax policy salience. Too often it is assumed that individuals are fully informed about federal tax policy and its incentives the moment a policy is implemented. Such a supposition may be grossly inaccurate. Chetty, Looney, and Kroft (2009) explored the impact of sales tax salience on consumption and found that demand for items reporting tax-inclusive prices dropped by 8 percent, suggesting that individuals do not fully optimize consumption by taking into account tax. Similarly, Finkelstein (2009) found that toll prices increased 20 to 40 percent with the adoption of electronic toll collection as salience diminished. Other studies exploring the impact of tax salience on consumption include Gallagher and Muehlegger (2008), Goldin and Homonoff (2013), and Taubinsky and Rees-Jones (2018).

Specifically, this research contributes to taxpayers' awareness of tax expenditures. In addition to the studies discussed earlier (Scholz 1994) and (Plueger 2009), Chetty and Saez (2013) explored the impact of providing program information to EITC recipients and found only weak evidence that information improvements affected EITC earnings behavior. Miller and Mumford (2015) explored the impact of complex tax changes on taxpayers' response to the Child and Dependent Care Credit (CDCC) expansion in 2003. They found that taxpayers increased child care expenditures in response to CDCC but ignored the complex interaction CDCC had with the Child Tax Credit (CTC) which may have resulted in suboptimal spending. Goldin and Listokin (2013) investigate taxpayer perceptions around the Charitable Deduction and Home Mortgage Interest Deduction. They found the lack of awareness pervasive among taxpayers with respect to both subsidies and tending toward lower income households. My investigation shows, similarly, higher income taxpayers are more apt to be aware and participate, along with filing jointly and using a paid preparer. Beyond previous research, my study exposes the spatial dynamic underpinning tax salience.

I also contribute to the sociological field of knowledge diffusion through social networks by documenting the diffusion and velocity of Saver's Credit behavior. Because I observe filing behavior spatio-temporally, I capture the magnitude and neighborhood clustering of knowledge at the time of the credit's initial availability and observe its diffusion over time.

Where research into salience suggests only some behavioral change through information, the field of social network research suggests that knowledge diffusion through social interaction may be more impactful. Three leading theories emerge from this line of research; weak ties (Granovetter 1973), structural holes (Burt 1992), and embeddedness (Polanyi 1968; Granovetter 1985). The theory of weak ties suggests that strong associational ties between individuals results in high information overlap while weak ties allow for the transfer of new information. Similarly, the structural holes theory says that innovation and new knowledge are produced by bridging gaps in information between dense clusters of closely associated individuals. Embeddedness, an expansion of the weak ties theory, argues that 'rational'[5] economic activity is embedded in pre-existing social attachments. In other words, the market economy is not independent from social relationships as would be assumed by neoclassical economics but constrained by and embedded in social networks. In all three theories, acquiring information is done through movement outside, or between, highly associated groups of individuals. As such, they all highlight the importance of social relationship for knowledge diffusion. In the case of tax policy, weak associations may lead to faster knowledge dispersion and more behavioral adjustment. By assuming residence serves as a weak social tie, my research contributes to this field by mapping the initial epicenters of information through participation and the radial spread of information across the Southeast region. It further empirically measures the strength of spatial spillovers over time.

This research also contributes to the field of peer effects by exploring if one's geographic location amplifies or dampens participation in the Saver's Credit. By looking at cross-neighborhood variation, which can be considered a weak social tie, and exploring

[5] By way of neoclassical rational-agent models

the spatial autocorrelation associated therein, I address the peer effects that one's choice of residential location has on participation. In 2003, Duflo and Saez created an experiment to test the dispersion of information and participation decisions around retirement savings vehicles at one university. Their research found strong spillover effects on information gathering and subsequent enrollment. In a more structured study examining how peer interactions affect financial investment decisions, Bursztyn et al. (2014), through a field experiment in Brazil looking at individuals' response to information pertaining to their peer's purchasing decision of a high-stakes financial asset, found that peer effects had a significant role in investment behavior, more so than Duflo and Saez (2003) suggested.

Beshears et al. (2015) also documented the peer effects on savings behavior but found contradictory results. They developed a field experiment to uncover the effects of disseminating peer savings information at one firm on zero and low percentage saving employees. It was initially assumed that presenting employees with peer savings participation rates would encourage nonsavers[6] and low-savers to increase participation. They found instead that nonsavers presented with peer information responded negatively, significantly reducing their likelihood to enroll. They suggest that this was in part caused by a sense of discouragement at observing higher achieved savings rates by others that they felt they could not achieve themselves.

---

[6] Here nonsavers are those individuals who were never enrolled in a 401(k) unless choosing to opt-in, as opposed to individuals who were initially automatically enrolled at a 6% contribution rate who subsequently opted-out.

I contribute to these fields by exploring the link between taxpayers at the neighborhood level to the Saver's Credit incentive to increase retirement savings contributions. Where Bursztyn et al. (2014) examined peer effects through very close social ties, Beshears et al. (2015) examined peer effects where the ties were much more arm's length. In my study, I examine peer effects at a similarly distant tie but rather than a connection through employment, I use a connection through residence.

Additionally, my research contributes to the growing literature studying the peer effects on tax reporting. Alm, Bloomquist and McKee (2013) explored the impact of peer tax compliance information on individual compliance. They found that providing any information on peer compliance had a significant and positive impact on filing and reporting decisions, but more information did not always improve compliance. Fortin, Lacroix, and Villeval (2007) examined the impact of peer effects on tax evasion. Contrary to expectations, they found that individual tax evasion was anti-conforming to peer tax evasion expectations.

While these studies examined the peer effects on general tax compliance through directed information transfer, my research focuses on tax expenditures and on passive observation of knowledge transfer. Not much work has been conducted in this field, with the notable exception of Chetty, Friedman, and Saez (2013). Here, they explored the peer effects of the EITC. In this study Chetty et al. use filing location to observe knowledge diffusion of the work incentives surrounding the EITC. By observing the labor market and EITC filing behavior of individuals moving between neighborhoods, they were able to witness increases in both as individuals moved from an area with low take-up rates to areas with high take-up rates. Further, they observed little to no declines as individuals moved away from high take-up rate areas to low take-up rate areas.

My research builds off these fields to apply such insights to the observed patterns of participation in the Saver's Credit. Where originally research assumed either an instantaneous awareness of the Saver's Credit or only a 5-year learning window, to include a near full understanding of its incentive structure, I relax those assumptions. In Chapter 4, I map the spread of participation over time as it radiates away from initially highly concentrated areas, showing that participation did not follow a homogenous growth trajectory. In Chapter 5, I empirically prove Chapter 4's observational conclusions and find that participation has spatial dependence. In Chapter 6, I first prove that traditional regression modeling is ill-equipped to handle spatial dependence when measuring determinants of participation. I therefore employ more appropriate spatial regression models to estimate the relationship between participation and filing characteristics, to include the size and direction of spatial spillovers. Finally, in Chapter 7, I measure the longer-run impact of the Saver's Credit, after allowing time for the heterogeneous growth of participation to occur.

The remainder of this study is outlined as follows. Before moving into each of these four parts, I formally outline the structure of the credit and review the literature exploring the Saver's Credit and its impact. Chapter 3 describes my data and provides some general descriptives. I then break out into my four-part study in Chapters 4, 5, 6, and 7. Chapter 8 provides concluding remarks.

CHAPTER 2 - REVIEW OF THE SAVER'S CREDIT

As described earlier, the Saver's Credit is a nonrefundable tax credit offered to low and middle-income households to encourage contributions into any qualified retirement plan[7]. The credit provides a 10, 20, or 50 percent match to contributions up to $2,000 per person, per tax unit that can be used only to offset tax liability[8]. The exact credit rate is determined by filing status and reported AGI, with 50 percent at the bottom of the distribution and 10 percent at the top of the capped distribution. To see how this works, take Tax Year 2015 as an example, a single filer with income below $18,250 could receive a 50 percent credit on contributions up to $2,000 to lower her tax liability, income between $18,251 and $19,750 could receive a 20 percent credit, income between $19,751 and $30,500 could receive a 10 percent credit, and income above $30,500 would make her ineligible. The exact threshold cutoffs for the various filing statuses are interrelated, with head of household thresholds equaling three-fourths that of married filing jointly and all other filing statuses equaling one-half of married filing jointly. Since the credit is offered at the person-level, a married filing jointly household could receive a credit for contributions up to $4,000.

In effect the credit offers a government match to contributions equal to 100 percent, 25 percent, or 11 percent, for 50, 20, or 10 percent credit rates respectively. To see this, suppose a single filer facing a 50 percent credit rate contributes $2,000 to a

[7] Qualified retirement plans include any type of IRA plans or DC plans.
[8] For primary or secondary filers only.

retirement plan. The filer should expect to receive a $1,000 tax credit, lowering her out of pocket cost of $2,000 down to only $1,000, equal to the amount of the credit. Doing this same exercise for the 20 percent and 10 percent credit results in a government match of 25 percent and 11 percent respectively. Keeping in mind this match rate can only be used to offset tax liability. While a taxpayer could conceivably receive a 100 match, if their tax liability is small, or zero, the realized match rate is diminished.

Two less apparent advantageous features of the credit are its independence from employer-sponsored plans and the income threshold computation. First, the credit is available for any contributions up to the $2,000 cap into any qualified retirement plan, to include Defined Contribution (DC) plans and Individual Retirement Accounts (IRAs), without regard to the employer-match. This offers the benefit of supporting employer-sponsored plans by further incentivizing individuals to contribute, taking advantage of both employer-sponsored benefits and government matching benefits.

The second advantage relates to how income is defined for rate threshold qualification. The income cutoffs for qualification are based on one's computed AGI for which retirement contributions, in the case of pre-tax programs, have already been removed. This is unlike the IRA deduction which requires computation of a modified AGI (MAGI) that adds back IRA contributions. This results in improving the likelihood of qualifying for the credit two-fold, first by making contributions and second by lowering one's AGI.

While these advantages serve to make the credit more enticing, earlier studies examining its impact have all consistently decided it has had both little utilization and little impact on savings behavior. Gale, Iwry, and Orszag (2005) lay out some problematic design features that serve only to limit the credit's effectiveness, they report that in the first two years of its offering about 5.4 million households claimed the credit. Comparing that to the 89.3 million tax returns filed in the same two-year period that had income below the credit cutoff and having a tax liability, such uptake suggests a lackluster reception. However, they also report that 71 percent of 401(k) plan sponsors believed that the credit had increased participation and a spokesperson for H&R Block stated that many first-time claimants were also first-time retirement contributors. Shortly thereafter, Koenig and Harvey (2005) estimated that nearly half of recipients were unable to receive the full credit due to its nonrefundability. Over a third of households with income below the credit threshold failed to claim the credit, averaging $184 of unclaimed credit per filer. They also reported nearly 68 percent of claimants only qualified for the lowest credit bracket of 10 percent, suggesting a significantly lower take up rate among lower income households.

To date, only three studies have examined the direct effects of the Saver's Credit on retirement contributions. In the first, Duflo et al. (2007) examined participation rates using H&R Block national retail tax return data for Tax Year 2005. They compared eligible tax units, defined as those who had tax liability, to ineligible tax units, or those with no liability. They found that the credit had, at best, a modest impact on retirement contribution decisions. They went on to offer inappropriate framing as a potential rationale for this low response. They point out that the Saver's Credit was perceived as a tax "credit" that only works to lower tax liability as opposed to a match on savings.

While informative, this study was conducted only a few years after the credit's introduction, limited to only one year of filing behavior, and focused exclusively on filers utilizing H&R Block's X-IRA program, making extrapolation across the full population of filers over a long period of time difficult.

In another study, Ramnath (2013) found the credit failed to generate any significant effect on overall retirement savings contributions. Upon using the 2002-2006 Individual Income Tax Cross Section Public Use Files (PUF), she observed considerable bunching at each of the credit brackets. This suggested some understanding by tax filers around the incentives created by the credit. She found that rather than increasing savings contributions, which she estimated to be between a $356 decline and $127 improvement, claimants on average lowered their AGI by as much as $1,400 to qualify for a higher credit, lowering tax liability by $400. Under the backdrop of this analysis, Ramnath also examined the drivers to credit participation. She found married filing jointly, self-employment, e-filing, and using a paid preparer all positively impacted participation. Though slightly more expansive in its window of years to previous studies, this research suffered from a small sample size of lower income households and severely limited information on retirement contributions reported on a tax return.

The last investigation was conducted by Heim and Lurie (2014). They used IRS's 1987 Family Panel to explore the impact of the Saver's Credit between filers having transitorily or permanently low income. They employed a difference-in-difference model comparing tax filers from 2000-2001, just before the Saver's Credit was enacted, to 2005-2006, a few years after its introduction. This allowed for individuals to both learn and respond to the incentives of the credit. They concluded that a large proportion of the credit's benefits were focused on individuals with only transitorily low-income levels, with more than 20 percent of the benefits going to these individuals. They also estimated the credit produced an overall 2.8 percentage point increase in the probability to contribute to a retirement program, with an estimated increase of $317 of contributions, with a standard deviation of $148. This estimate is far greater than previous research, calling into question earlier findings, is dependent on a rather small sampling rate for lower income households and is limited to only the first few years of credit availability.

Now that nearly two decades have passed, several questions remain unanswered. Where earlier studies focused on trying to quantify the impact of the Saver's Credit on retirement contributions with mixed results, few explored the more basic question of what influences participation. Those that did either ignored, or were unable to account for, the role that spatial interactions play on participation. Further, none of these studies explored behavior after 2006, thus not account for the slow, but steady, expansion of credit utilization, the impact of credit indexing, or the hit of the Great Recession. This research serves to shed light on these remaining questions to answer where participation occurred, when it was occurring, how it has changed over time, quantifying the simultaneous spatial dependence that underpins observed patterns, and the longer run impact of the credit after allowing for participation growth to occur. Through this research I answer these questions and contribute to the broader discussion around tax policy salience, impacts, and relationships to taxpayer behavior.

CHAPTER 3 - DATA

I draw on the population of individual income tax filers spanning tax years 1999 through 2013 compiled and warehoused at the Internal Revenue Service (IRS). Starting in 1999 affords up to three years of information prior to the introduction of the Saver's Credit in 2002, allowing for a strong base on which to compare post-credit behavior. I stop at 2013 to exclude any late-filer interactions.[9] Among the advantages of this dataset is the ability to observe savings behavior, credit eligibility and participation, both longitudinally and at high spatial granularity.

The dataset I produce required combining tax return information with information returns and Social Security demographics. From the tax return, I observe a whole host of pertinent characteristics including filing status, AGI, tax liability, and filing location. In addition to these, I can also observe a range of other useful information including wages, self-employment income, and whether someone claimed the Saver's Credit, along with the amount.

I merge in information returns filed by third parties on the taxpayer's behalf. Such information includes Form W-2 and Form 5498. These provide information for annual retirement contributions to all the various qualified retirement plans. Form W-2 reports, along with wages, elective deferrals into various qualified retirement plans, particularly 401(k)-type plans, Savings Incentive Match Plan for Employees (SIMPLE), and

[9] At the time of this investigation, tax year 2016 was just kicking off and previous research into tax return filing patterns have found approximately 3 percent of tax returns for any calendar year are filed up to 2 years later (Weber, Bryant, 2005).

Table 3.1 Number of Taxpayers by Credit Eligibility, 1999-2013

(Counts in thousands)

Tax Year	Population Count [1]	Eligiblity Group [2]	
		Subgroup 1	Subgroup 2
1999	111,409	40,849	8,146
2000	118,450	49,293	7,738
2001	119,708	46,789	n.a.
2002	120,033	44,681	9,806
2003	119,263	41,835	8,871
2004	117,837	39,189	8,441
2005	116,619	36,343	8,084
2006	115,223	33,068	7,504
2007	117,353	32,970	8,052
2008	112,666	31,838	7,821
2009	111,906	33,634	8,188
2010	110,410	33,176	7,527
2011	107,901	32,070	7,468
2012	103,939	30,893	7,296
2013	100,549	30,456	7,412

[1] Population reflects tax filers between 18 and 57 in 1999 with income below $100,000

[2] Subgroup 1 includes tax filers with AGI below threshold and a tax liability.

Sugroup 2 includes tax filers who have AGI below threshold, a tax liability, and making a retirement contribution.

n.a. - Not Available

Source: Compliance Datawarehouse, IRS

Simplified Employee Pension (SEP) plans. Form 5498 reports individuals’ contributions to IRAs. I then add limited demographic information reported by Social Security including age and gender.

Because this credit attempts to target lower income households prior to retirement, I limit the population to individuals of prime working age. Specifically, I only include those between the ages of 18 and 57 in the base year and still alive in 2013. This ensures I exclude anyone who turns 70.5 by 2013 and hit IRA Required Minimum Distribution (RMD) requirements. Additionally, I remove all returns with income above $100,000, since they would be ineligible for the credit. Table 3.1 reports the breakdown of my population by tax year.

Given the structure of the credit, the definition of credit eligibility can be determined in one of a few ways, each of which could change the interpretation of credit success. I have thus subdivided my initial population into more targeted eligibility definitions. The first, termed Subgroup 1, limits the population to those who's income was below the credit qualifying threshold and had a tax liability. Here, I compute tax liability to be the remaining liability after other tax credits have been applied. Subgroup 2 is defined as those tax filers who, in addition to having income below the threshold and having tax liability, also have contributed to a qualified retirement plan. This last population meets all filing, income, and saving requirements to claim the credit. In theory, this population should show 100 percent participation and it is this population that is of particular interest in the tax salience field of research. See Table 3.2 for a population break down of demographics by the various eligibility definitions.

While the data are rather exhaustive, there are two limitations that should be highlighted. As mentioned, W-2 data reports DC plans; however, for 2001 exclusively this information was irrevocably omitted from the original data source. In other words, retirement contribution information for that year includes only contributions to IRA type plans and no contributions to DC plans. In the data this represents a 66 percent drop in the number of tax filers reportedly contribution to a retirement plan from the previous year. Going forward, I simply omit 2001 from all analysis.

The second limitation is the filing location reported on the tax return. While in most cases households report their home address, there is no legal requirement to do so. As such some returns report locations other than their legal place of residence, such as a business address, paid preparer address, or P.O. Box. This may adversely affect insights gleaned from mapping the distribution of credit knowledge and dispersion of knowledge over time if there are large differences between residence and tax reported address. There is currently very little research into the extent of this issue. Larrimore, Mortenson, and Splinter (2017) in trying to create household units from tax units, come the closest to addressing this concern. They report that tax units generally fall within 4 percent (+/-) of Census reported household populations at the state level, with the major exception of Alaska at nearly 5 percent higher. To address this limitation, I focus exclusively on the continental U.S. Moreover, in most instances I further narrow focus to the Southeast region for computational and graphical density regions.

Because the underlying data is not a sample but the population of tax filers, I can be quite precise in my analysis when exploring spatial impact. However, there is one additional data limitation that arises when mapping addresses: ZIP Codes are a U.S. Postal Service creation to assist in the delivery of post, not a geographic locator. Thus, not all ZIP Codes map to land areas; some ZIP Codes map only to a building within a land area. One example of this would be the State Department; this District of Columbia based building has its own ZIP Code from which tax information theoretically could be filed. I drop all such ZIP Codes from my analysis because filing activity from these locations often do not reflect the behavior of filing for the surrounding land mass areas nor is there a geographic area to map to.

Table 3.2 Number of Taxpayers by Credit Eligibility, 2002-2013

(Counts in thousands)

Demographics	2002 Population Count	2002 Claiming Saver's Credit	2003 Population Count	2003 Claiming Saver's Credit	2004 Population Count	2004 Claiming Saver's Credit
Full Population [1]						
Population Size	120,033	7,064	119,263	7,193	117,837	7,049
Making retirement contribution	36,353	4,977	34,980	5,025	35,224	4,972
Using outside preparer	48,309	3,391	50,635	3,480	50,356	3,361
Using a paid preparer	47,374	3,308	50,511	3,474	50,242	3,355
Single	42,347	1,072	40,512	1,090	38,912	1,080
Married filing Jointly	60,467	4,655	61,265	4,746	61,116	4,575
Head of Household	17,220	1,337	17,485	1,357	17,810	1,393
Age <35	50,812	2,651	48,004	2,581	45,083	2,487
Age 35<45	33,366	2,139	32,759	2,112	31,982	1,986
Age 45<55	27,744	1,748	28,122	1,817	28,256	1,787
Age 55 and older	8,112	525	10,378	683	12,517	788
SE income 90% or more [2]	7,709	209	8,014	221	8,132	219
SE income 50% or more [2]	9,929	361	10,305	378	10,525	372
Wage income 90% or more	70,762	4,007	68,806	4,074	67,355	4,074
Wage income 50% or more	90,878	5,264	88,800	5,322	86,964	5,238
Qualified for 50% Credit rate	41,491	1,558	39,854	1,488	37,265	1,417
Qualified for 20% credit rate	3,719	500	3,689	506	3,584	499
Qualified for 10% credit rate	24,321	5,006	24,278	5,198	23,850	5,132
Zero to three exemptions	65,994	4,275	65,256	4,328	63,976	4,179
Four or more exemptions	23,471	1,447	23,622	1,465	23,690	1,485
Male	58,142	3,090	57,676	3,163	56,922	3,091

Footnotes at end of table

Table 3.2 Number of Taxpayers by Credit Eligibility, 2002-2013, cont.

Demographics	2002 Population Count	2002 Claiming Saver's Credit	2003 Population Count	2003 Claiming Saver's Credit	2004 Population Count	2004 Claiming Saver's Credit
Subgroup 1: AGI below Saver's Credit threshold and Tax Liability						
Population Size	44,681	7,030	41,835	7,148	39,189	6,992
Making retirement contribution	9,806	4,971	8,871	5,016	8,441	4,961
Using outside preparer	18,019	3,391	17,705	3,479	16,800	3,360
Using a paid preparer	17,507	3,308	17,642	3,473	16,742	3,354
Single	16,387	1,070	14,856	1,086	13,351	1,076
Married filing Jointly	20,509	4,626	19,395	4,708	18,385	4,529
Head of Household	7,784	1,334	7,584	1,353	7,453	1,388
Age <35	22,212	2,629	19,700	2,553	17,548	2,455
Age 35<45	10,933	2,131	10,183	2,100	9,449	1,971
Age 45<55	8,766	1,745	8,536	1,812	8,233	1,780
Age 55 and older	2,770	525	3,417	682	3,959	787
SE income 90% or more [2]	1,941	209	1,929	221	1,964	219
SE income 50% or more [2]	2,792	361	2,752	378	2,780	372
Wage income 90% or more	27,717	4,000	25,605	4,062	23,979	4,058
Wage income 50% or more	34,369	5,254	31,860	5,306	29,714	5,218
Qualified for 50% Credit rate	17,959	1,545	15,657	1,473	13,872	1,399
Qualified for 20% credit rate	3,382	497	3,260	501	3,095	494
Qualified for 10% credit rate	23,340	4,988	22,918	5,173	22,222	5,099
Zero to three dependents	23,673	4,255	21,936	4,301	20,236	4,147
Four or more dependents	9,433	1,439	8,815	1,453	8,469	1,471
Male	21,485	3,079	20,094	3,148	18,778	3,071

Footnotes at end of table

Table 3.2 Number of Taxpayers by Credit Eligibility, 2002-2013, cont.

Demographics	2002		2003		2004	
	Population Count	Claiming Saver's Credit	Population Count	Claiming Saver's Credit	Population Count	Claiming Saver's Credit
Subgroup 2: AGI below Saver's Credit threshold and Tax Liability and Retirement Contribution						
Population Size	9,806	4,971	8,871	5,016	8,441	4,961
Making retirement contribution	9,806	4,971	8,871	5,016	8,441	4,961
Using outside preparer	4,070	2,385	3,848	2,434	3,670	2,375
Using a paid preparer	3,965	2,323	3,838	2,428	3,662	2,370
Single	2,601	1,002	2,276	1,008	2,144	1,006
Married filing Jointly	4,923	2,712	4,488	2,738	4,212	2,644
Head of Household	2,283	1,257	2,107	1,269	2,086	1,310
Age <35	3,804	1,850	3,272	1,785	3,064	1,742
Age 35<45	2,814	1,507	2,472	1,475	2,255	1,398
Age 45<55	2,421	1,239	2,233	1,274	2,114	1,262
Age 55 and older	767	375	894	482	1,008	559
SE income 90% or more [2]	299	153	289	161	290	163
SE income 50% or more [2]	475	252	455	261	453	262
Wage income 90% or more	6,955	3,612	6,261	3,656	6,017	3,680
Wage income 50% or more	8,767	4,479	7,892	4,512	7,521	4,482
Qualified for 50% Credit rate	2,381	1,109	2,036	1,054	1,873	1,010
Qualified for 20% credit rate	694	355	630	356	602	354
Qualified for 10% credit rate	6,731	3,508	6,205	3,606	5,967	3,596
Zero to three dependents	5,563	2,954	5,000	2,958	4,689	2,879
Four or more dependents	1,861	986	1,691	993	1,663	1,019
Male	4,671	2,373	4,214	2,403	3,955	2,350

Footnotes at end of table

Table 3.2 Number of Taxpayers by Credit Eligibility, 2002-2013, cont.

Demographics	2005		2006		2007	
	Population Count	Claiming Saver's Credit	Population Count	Claiming Saver's Credit	Population Count	Claiming Saver's Credit
Full Population [1]						
Population Size	116,619	6,845	115,223	6,620	117,353	7,193
Making retirement contribution	35,676	4,860	35,617	4,706	36,183	5,173
Using outside preparer	49,828	3,227	49,084	3,048	48,765	3,230
Using a paid preparer	49,745	3,223	48,994	3,045	48,645	3,225
Single	38,034	1,111	37,445	1,098	39,730	1,225
Married filing Jointly	60,763	4,381	59,753	4,149	59,458	4,434
Head of Household	17,822	1,353	18,024	1,374	18,165	1,534
Age <35	42,072	2,345	39,120	2,212	36,614	2,306
Age 35<45	31,429	1,870	30,760	1,761	30,739	1,890
Age 45<55	28,486	1,748	28,538	1,696	29,752	1,872
Age 55 and older	14,631	882	16,805	951	20,248	1,126
SE income 90% or more [2]	8,364	211	8,753	201	11,085	196
SE income 50% or more [2]	10,841	361	11,220	345	13,537	343
Wage income 90% or more	65,813	3,987	64,319	3,884	64,906	4,252
Wage income 50% or more	85,220	5,108	83,503	4,958	84,013	5,435
Qualified for 50% Credit rate	34,969	1,322	33,014	1,214	36,309	1,325
Qualified for 20% credit rate	3,483	481	3,317	465	3,826	589
Qualified for 10% credit rate	23,423	5,041	22,602	4,941	22,337	5,279
Zero to three exemptions	62,741	3,979	61,246	3,785	61,043	4,056
Four or more exemptions	23,785	1,470	23,851	1,473	24,238	1,660
Male	56,343	3,011	55,623	2,902	56,672	3,162

Footnotes at end of table

Table 3.2 Number of Taxpayers by Credit Eligibility, 2002-2013, cont.

Demographics	2005		2006		2007	
	Population Count	Claiming Saver's Credit	Population Count	Claiming Saver's Credit	Population Count	Claiming Saver's Credit
Subgroup 1: AGI below Saver's Credit threshold and Tax Liability						
Population Size	36,343	6,778	33,068	6,540	32,970	7,096
Making retirement contribution	8,084	4,846	7,504	4,687	8,052	5,146
Using outside preparer	15,733	3,226	14,469	3,048	14,395	3,230
Using a paid preparer	15,692	3,222	14,427	3,044	14,342	3,224
Single	11,920	1,105	10,413	1,091	10,063	1,217
Married filing Jointly	17,250	4,326	15,754	4,084	15,925	4,357
Head of Household	7,173	1,346	6,901	1,365	6,982	1,522
Age <35	15,218	2,308	12,765	2,171	11,550	2,259
Age 35<45	8,773	1,851	8,010	1,737	8,045	1,859
Age 45<55	7,928	1,740	7,509	1,685	7,831	1,858
Age 55 and older	4,424	879	4,784	948	5,543	1,120
SE income 90% or more [2]	1,924	211	1,818	201	1,750	196
SE income 50% or more [2]	2,715	361	2,562	344	2,495	343
Wage income 90% or more	22,058	3,968	19,803	3,859	19,353	4,219
Wage income 50% or more	27,333	5,084	24,597	4,926	24,273	5,393
Qualified for 50% Credit rate	12,002	1,304	10,178	1,195	9,916	1,302
Qualified for 20% credit rate	2,920	474	2,688	457	3,094	578
Qualified for 10% credit rate	21,421	5,000	20,201	4,888	19,960	5,216
Zero to three dependents	18,497	3,941	16,590	3,741	16,458	4,002
Four or more dependents	8,054	1,452	7,586	1,451	7,715	1,634
Male	17,359	2,986	15,722	2,872	15,667	3,125

Footnotes at end of table

Table 3.2 Number of Taxpayers by Credit Eligibility, 2002-2013, cont.

Demographics	2005 Population Count	2005 Claiming Saver's Credit	2006 Population Count	2006 Claiming Saver's Credit	2007 Population Count	2007 Claiming Saver's Credit
Subgroup 2: AGI below Saver's Credit threshold and Tax Liability and Retirement Contribution						
Population Size	8,084	4,846	7,504	4,687	8,052	5,146
Making retirement contribution	8,084	4,846	7,504	4,687	8,052	5,146
Using outside preparer	3,499	2,290	3,254	2,167	3,462	2,321
Using a paid preparer	3,493	2,287	3,248	2,164	3,454	2,316
Single	2,052	1,039	1,865	1,022	1,985	1,150
Married filing Jointly	3,990	2,535	3,642	2,378	3,870	2,552
Head of Household	2,041	1,273	1,996	1,286	2,196	1,444
Age <35	2,855	1,661	2,549	1,576	2,597	1,669
Age 35<45	2,099	1,322	1,914	1,246	2,045	1,354
Age 45<55	2,026	1,238	1,887	1,196	2,060	1,333
Age 55 and older	1,103	625	1,154	668	1,351	791
SE income 90% or more [2]	280	157	260	148	254	144
SE income 50% or more [2]	436	253	405	239	402	238
Wage income 90% or more	5,779	3,613	5,357	3,510	5,729	3,867
Wage income 50% or more	7,218	4,394	6,706	4,265	7,221	4,713
Qualified for 50% Credit rate	1,714	946	1,502	866	1,599	954
Qualified for 20% credit rate	575	342	536	330	670	422
Qualified for 10% credit rate	5,794	3,559	5,466	3,490	5,783	3,770
Zero to three dependents	4,414	2,750	4,029	2,614	4,276	2,825
Four or more dependents	1,651	1,015	1,613	1,019	1,812	1,169
Male	3,767	2,292	3,467	2,191	3,708	2,396

Footnotes at end of table

Table 3.2 Number of Taxpayers by Credit Eligibility, 2002-2013, cont.

Demographics	2008 Population Count	2008 Claiming Saver's Credit	2009 Population Count	2009 Claiming Saver's Credit	2010 Population Count	2010 Claiming Saver's Credit
Full Population [1]						
Population Size	112,666	7,279	111,906	7,834	110,410	7,362
Making retirement contribution	35,758	5,244	33,601	5,582	31,821	5,233
Using outside preparer	47,113	3,191	46,645	3,365	45,665	3,110
Using a paid preparer	46,987	3,185	46,524	3,359	45,568	3,104
Single	36,449	1,242	35,214	1,327	34,887	1,232
Married filing Jointly	58,437	4,474	59,163	4,983	58,053	4,658
Head of Household	17,779	1,563	17,529	1,525	17,470	1,471
Age <35	32,897	2,185	29,512	2,108	26,358	1,819
Age 35<45	29,349	1,897	28,965	2,035	28,490	1,905
Age 45<55	28,817	1,931	29,142	2,166	28,943	2,062
Age 55 and older	21,603	1,267	24,288	1,525	26,618	1,576
SE income 90% or more [2]	8,760	162	8,978	163	8,954	165
SE income 50% or more [2]	10,949	293	11,172	302	11,258	302
Wage income 90% or more	60,743	4,309	56,951	4,436	53,808	4,131
Wage income 50% or more	79,396	5,519	75,309	5,838	72,279	5,444
Qualified for 50% Credit rate	33,811	1,397	36,227	1,432	35,727	1,425
Qualified for 20% credit rate	2,768	431	2,779	416	3,065	479
Qualified for 10% credit rate	22,461	5,451	23,884	5,986	22,511	5,457
Zero to three exemptions	58,841	4,068	58,513	4,465	57,361	4,146
Four or more exemptions	23,912	1,722	23,833	1,743	23,698	1,694
Male	54,333	3,215	53,808	3,530	53,079	3,293

Footnotes at end of table

Table 3.2 Number of Taxpayers by Credit Eligibility, 2002-2013, cont.

Demographics	2008 Population Count	2008 Claiming Saver's Credit	2009 Population Count	2009 Claiming Saver's Credit	2010 Population Count	2010 Claiming Saver's Credit
Subgroup 1: AGI below Saver's Credit threshold and Tax Liability						
Population Size	31,838	7,168	33,634	7,718	33,176	7,240
Making retirement contribution	8,017	5,210	8,404	5,544	7,695	5,192
Using outside preparer	13,738	3,191	14,278	3,364	14,044	3,109
Using a paid preparer	13,680	3,184	14,220	3,358	13,996	3,103
Single	9,416	1,232	9,712	1,316	9,613	1,221
Married filing Jointly	15,686	4,389	17,444	4,893	17,173	4,565
Head of Household	6,736	1,548	6,478	1,509	6,390	1,455
Age <35	10,043	2,136	9,299	2,061	8,220	1,774
Age 35<45	7,762	1,859	8,205	1,994	8,079	1,861
Age 45<55	7,833	1,913	8,639	2,147	8,634	2,040
Age 55 and older	6,199	1,259	7,491	1,517	8,243	1,565
SE income 90% or more [2]	1,593	162	1,625	163	1,699	165
SE income 50% or more [2]	2,237	293	2,327	302	2,443	302
Wage income 90% or more	18,103	4,270	17,540	4,395	16,189	4,087
Wage income 50% or more	23,033	5,468	22,953	5,784	21,655	5,385
Qualified for 50% Credit rate	9,721	1,368	9,977	1,403	10,442	1,394
Qualified for 20% credit rate	2,223	423	2,233	409	2,469	469
Qualified for 10% credit rate	19,894	5,377	21,424	5,906	20,266	5,378
Zero to three dependents	15,869	4,007	17,166	4,400	16,766	4,078
Four or more dependents	7,574	1,692	7,558	1,714	7,636	1,664
Male	15,080	3,171	16,122	3,484	15,872	3,245

Footnotes at end of table

Table 3.2 Number of Taxpayers by Credit Eligibility, 2002-2013, cont.

Demographics	2008 Population Count	2008 Claiming Saver's Credit	2009 Population Count	2009 Claiming Saver's Credit	2010 Population Count	2010 Claiming Saver's Credit
Subgroup 2: AGI below Saver's Credit threshold and Tax Liability and Retirement Contribution						
Population Size	8,017	5,210	8,404	5,544	7,695	5,192
Making retirement contribution	8,017	5,210	8,404	5,544	7,695	5,192
Using outside preparer	3,389	2,295	3,491	2,392	3,173	2,205
Using a paid preparer	3,380	2,290	3,481	2,387	3,166	2,201
Single	1,947	1,169	2,014	1,247	1,797	1,154
Married filing Jointly	3,864	2,563	4,279	2,859	3,894	2,647
Head of Household	2,206	1,478	2,111	1,438	2,004	1,391
Age <35	2,406	1,590	2,241	1,514	1,890	1,300
Age 35<45	2,024	1,363	2,106	1,440	1,928	1,348
Age 45<55	2,091	1,376	2,281	1,532	2,103	1,459
Age 55 and older	1,496	882	1,776	1,057	1,774	1,086
SE income 90% or more [2]	211	116	202	116	196	116
SE income 50% or more [2]	338	197	334	202	323	201
Wage income 90% or more	5,692	3,938	5,648	4,033	5,132	3,762
Wage income 50% or more	7,225	4,813	7,407	5,058	6,780	4,741
Qualified for 50% Credit rate	1,644	1,004	1,641	1,016	1,580	1,005
Qualified for 20% credit rate	500	319	500	323	509	338
Qualified for 10% credit rate	5,873	3,887	6,263	4,205	5,606	3,849
Zero to three dependents	4,225	2,829	4,524	3,064	4,108	2,835
Four or more dependents	1,865	1,222	1,838	1,221	1,754	1,187
Male	3,695	2,426	3,944	2,625	3,508	2,397

Footnotes at end of table

Table 3.2 Number of Taxpayers by Credit Eligibility, 2002-2013, cont.

Demographics	2011 Population Count	2011 Claiming Saver's Credit	2012 Population Count	2012 Claiming Saver's Credit	2013 Population Count	2013 Claiming Saver's Credit
Full Population [1]						
Population Size	107,901	7,340	103,939	7,443	100,549	7,622
Making retirement contribution	31,085	5,225	30,144	5,259	29,256	5,381
Using outside preparer	43,999	3,021	42,518	3,010	40,958	3,034
Using a paid preparer	43,932	3,017	42,473	3,007	40,958	3,034
Single	34,591	1,240	32,775	1,271	31,643	1,319
Married filing Jointly	56,302	4,635	54,744	4,641	53,058	4,746
Head of Household	17,008	1,466	16,419	1,531	15,849	1,557
Age <35	22,946	1,645	19,266	1,496	15,810	1,331
Age 35<45	27,844	1,924	27,058	2,013	26,351	2,114
Age 45<55	28,296	2,074	27,143	2,110	26,187	2,185
Age 55 and older	28,815	1,698	30,472	1,824	32,201	1,993
SE income 90% or more [2]	9,145	168	8,731	168	8,485	179
SE income 50% or more [2]	11,412	305	10,956	306	10,718	321
Wage income 90% or more	52,897	4,177	50,485	4,279	48,501	4,389
Wage income 50% or more	70,369	5,451	67,158	5,544	64,344	5,676
Qualified for 50% Credit rate	35,524	1,439	34,053	1,455	33,635	1,537
Qualified for 20% credit rate	2,607	410	2,742	450	2,606	452
Qualified for 10% credit rate	22,254	5,492	21,411	5,538	20,877	5,633
Zero to three exemptions	55,695	4,107	53,396	4,091	51,302	4,108
Four or more exemptions	23,376	1,730	22,857	1,859	22,477	1,997
Male	51,800	3,293	49,681	3,336	47,929	3,426

Footnotes at end of table

Table 3.2 Number of Taxpayers by Credit Eligibility, 2002-2013, cont.

Demographics	2011 Population Count	2011 Claiming Saver's Credit	2012 Population Count	2012 Claiming Saver's Credit	2013 Population Count	2013 Claiming Saver's Credit
Subgroup 1: AGI below Saver's Credit threshold and Tax Liability						
Population Size	32,070	7,207	30,893	7,226	30,456	7,443
Making retirement contribution	7,630	5,177	7,491	5,144	7,602	5,312
Using outside preparer	13,402	3,020	13,084	2,987	12,832	3,034
Using a paid preparer	13,369	3,017	13,063	2,985	12,832	3,033
Single	9,218	1,226	8,443	1,194	8,349	1,301
Married filing Jointly	16,776	4,533	16,501	4,523	16,392	4,611
Head of Household	6,076	1,447	5,949	1,509	5,715	1,532
Age <35	7,001	1,601	5,818	1,437	4,816	1,287
Age 35<45	7,814	1,874	7,623	1,938	7,540	2,041
Age 45<55	8,392	2,048	8,089	2,060	7,985	2,144
Age 55 and older	8,863	1,685	9,363	1,792	10,115	1,972
SE income 90% or more [2]	1,718	168	1,700	167	1,669	179
SE income 50% or more [2]	2,435	305	2,403	304	2,365	321
Wage income 90% or more	15,931	4,126	15,327	4,163	15,118	4,316
Wage income 50% or more	21,075	5,384	20,229	5,404	19,869	5,579
Qualified for 50% Credit rate	10,069	1,405	9,619	1,416	9,511	1,491
Qualified for 20% credit rate	2,087	400	2,220	439	2,115	440
Qualified for 10% credit rate	19,914	5,402	19,054	5,372	18,829	5,512
Zero to three dependents	16,195	4,034	15,526	3,985	15,147	4,013
Four or more dependents	7,457	1,698	7,484	1,816	7,546	1,953
Male	15,289	3,239	14,615	3,243	14,375	3,351

Footnotes at end of table

Table 3.2 Number of Taxpayers by Credit Eligibility, 2002-2013, cont.

	2011		2012		2013	
Demographics	Population Count	Claiming Saver's Credit	Population Count	Claiming Saver's Credit	Population Count	Claiming Saver's Credit
Subgroup 2: AGI below Saver's Credit threshold and Tax Liability and Retirement Contribution						
Population Size	7,630	5,177	7,491	5,144	7,602	5,312
Making retirement contribution	7,630	5,177	7,491	5,144	7,602	5,312
Using outside preparer	3,079	2,142	3,032	2,103	3,043	2,139
Using a paid preparer	3,074	2,139	3,029	2,101	3,043	2,138
Single	1,796	1,161	1,701	1,119	1,788	1,216
Married filing Jointly	3,852	2,630	3,785	2,597	3,821	2,648
Head of Household	1,982	1,386	2,005	1,428	1,993	1,448
Age <35	1,695	1,179	1,478	1,050	1,290	943
Age 35<45	1,939	1,364	1,964	1,404	2,033	1,487
Age 45<55	2,097	1,469	2,067	1,467	2,118	1,532
Age 55 and older	1,899	1,164	1,982	1,222	2,160	1,349
SE income 90% or more [2]	202	119	197	117	209	127
SE income 50% or more [2]	328	203	321	201	338	215
Wage income 90% or more	5,153	3,805	5,082	3,806	5,152	3,941
Wage income 50% or more	6,754	4,750	6,652	4,734	6,739	4,888
Qualified for 50% Credit rate	1,582	1,012	1,562	1,011	1,605	1,062
Qualified for 20% credit rate	445	297	492	332	484	333
Qualified for 10% credit rate	5,603	3,868	5,437	3,801	5,512	3,918
Zero to three dependents	4,058	2,810	3,944	2,755	3,933	2,783
Four or more dependents	1,785	1,214	1,863	1,291	1,962	1,389
Male	3,476	2,393	3,406	2,372	3,465	2,455

[1] Full Population reflects tax filers between 18 and 57 in 1999 with income below $100,000

[2] SE is Self-Employment

Source: author's computation; Compliance Datawarehouse, IRS

CHAPTER 4 - THE GEOGRAPHIC PREVALENCE OF THE SAVER'S CREDIT PARTICIPATION

This first study serves to shed light on the spatial variability underlying aggregated national statistics and to highlight the disparate shifts over time in the spatial concentrations. The purpose being to observe the progression of awareness through the varying distributional shifts in participation in the Saver's Credit over time. Further, it explores the overlap of movement between participation in the Saver's Credit and participation in retirement savings programs through contributions to glean whether such activities appear correlated.

As a first step, if the Saver's Credit is growing in awareness and thus significance, I would expect to see improvement over time in participation across all of my populations of interest, the full population and Subgroups 1 and 2. This is exactly what I find. Figure 4.1 represents the distribution of filers across all 14 years subdivided by claiming behavior.[10] As the definition of eligibility becomes more targeted, the percentage of filers claiming the credit improves. Unsurprisingly, the first year of credit availability shows lower participation than subsequent years. Recall that Subgroup 1 is the target group for behavioral change. We observe that by 2013 this subgroup shows nearly a quarter of filers claiming the credit. This would suggest that the credit is impacting filers, at least by improving their awareness. Subgroup 2, again, ought to be nearly 100 percent as they meet all credit requirements. However, what we see is that in the first year only

[10] Because my defined population is a subset of the filing population as a whole, the average number of filers claiming the Saver's Credit is slightly higher than that reported in Table 1.2.

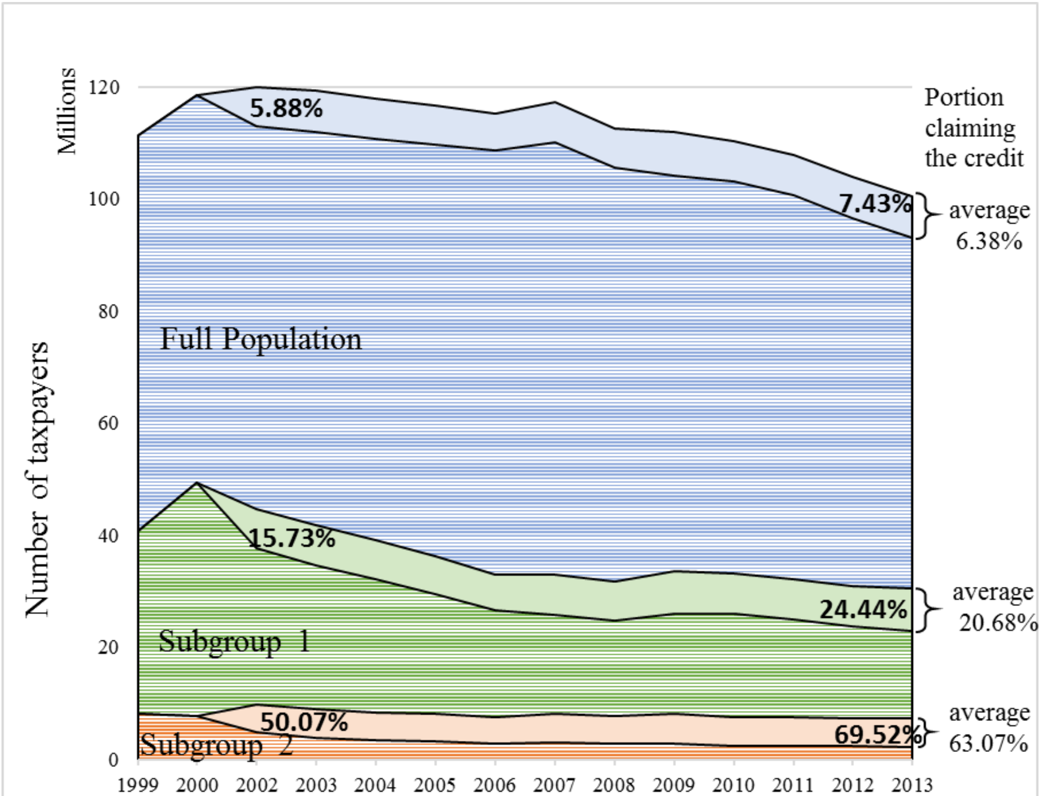


Figure 4.1: Distribution of Taxpayers by Eligibility Group and Credit Participation, 1999-2013

half of filers claimed the credit and by 2013 just under 70 percent claimed. These observations beg the question of how local jurisdictions fare, what patterns, if any, exist in the expansion of take-up, and whether more should be done to make eligible filers aware.

This study is broken out into four subsections. The next section presents national density distributions. Section 4.1 looks at the spatial distributions between Saver's Credit participation and retirement contribution participation for the Southeast region over time.

Section 4.2 overlays these two behaviors to draw insight into correlation. And Section 4.3 offers interpretations and discussion.

Section 4.1 - Density Distribution

Here, I examine the distribution of credit participation compared to participation in retirement programs through contributions over time across 5-digit ZIP Codes. Figures 4.2A through C show the density distribution for the percent of filers across ZIP Codes making a retirement contribution (left panel) and claiming the Saver's Credit (right panel). For the whole population, as highlighted above, we again see that on average the participation rates per ZIP Code, right panel, improved as time passed. What we also see is that while there is a shift in claiming activity over time, there is also a slight convergence across ZIP Codes. This is indicated by the narrowing and increased spike in the distribution over time. This suggests that some areas are improving participation but many are not. As compared to the pattern in retirement contribution behavior, on the left, we observe that just after the introduction of the credit there was a small increase in the percent of filers. However, there are indications that this may have been a trend starting prior to the offering of the Saver's Credit.

Again, because this credit is targeted at lower income households, it is more useful to examine these trends for the subpopulations. Figure 4.2B reports the distribution for Subgroup 1. The right panel repeats the trend from above with a shifting of concentration to higher percentages claiming the credit. The shift is more pronounced than before, suggesting that the target population is increasingly taking advantage. Looking at the distribution of retirement contributions, left panel, we actually see a

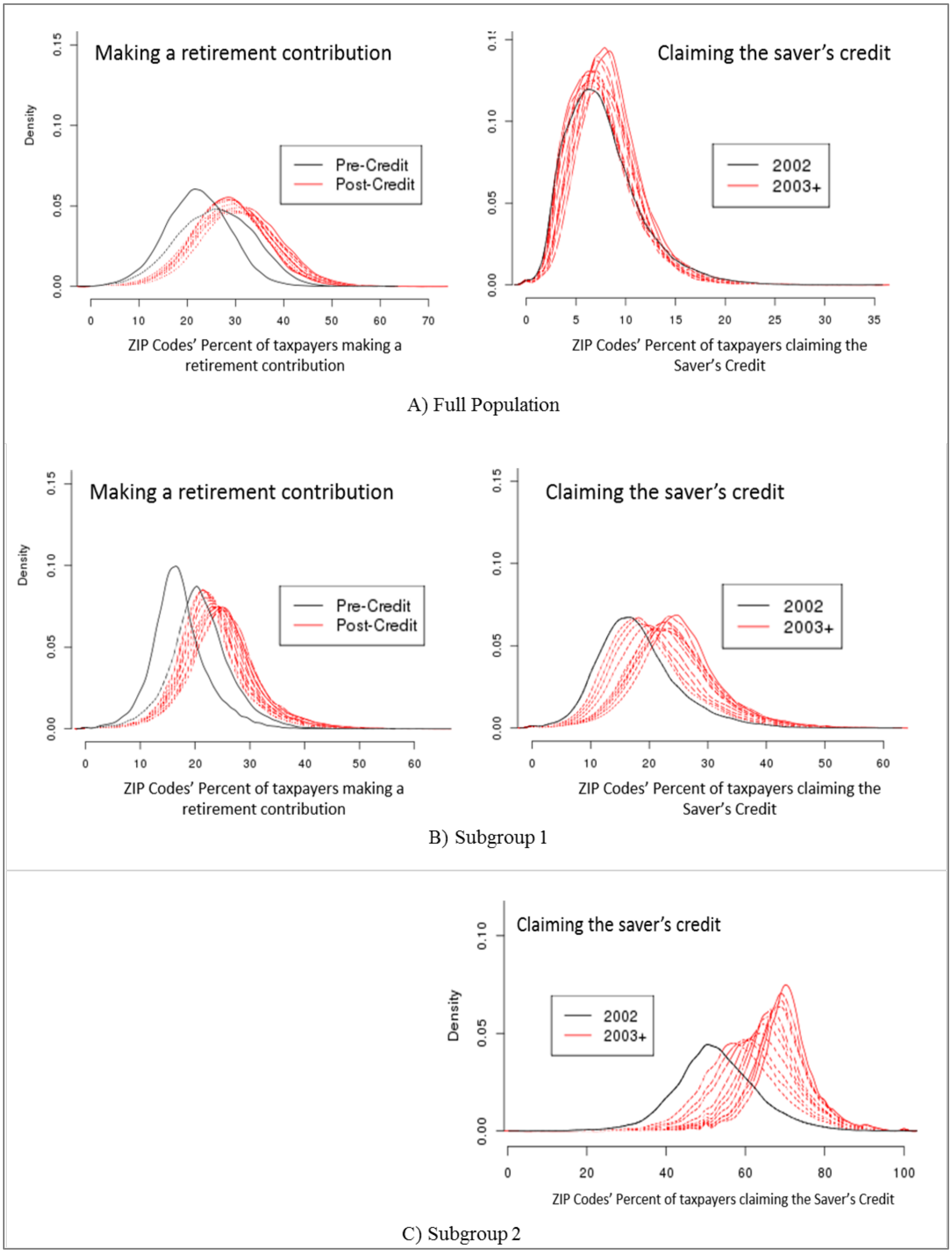


Figure 4.2: ZIP Code Distribution of the Percent of Taxpayers, 1999-2013

widening of the distribution after 2002, indicating that while more areas overall increased their contribution behavior the variance between ZIP Codes slightly grew as well. In combination, there are large positive shifts in both credit activity and contributions compared to the full population. This may suggest some causal relationship. That said, the only definitive conclusion is that there is a positive correlation, but it is unclear from this which behavior is driving the other.

Figure 4.2C presents the density distribution for Subgroup 2. Here, by definition 100 percent of all areas have retirement contributions, and correspondingly, we should expect to see nearly 100 percent claiming behavior across all areas as well. However, this is not what we find. Instead, we see that the distribution of participation across areas is shifting upward every year but never peaks above 70 percent. This movement right confirms that tax filers are improving their participation and suggests a growth in awareness each year. The distance between density plots provides evidence for that growth. Between 2002 and 2003 there appears quite a large shift in participation, not seen by any other year-to-year comparison. While each year shows improvement over the last, the distribution is narrowing, indicating that the ZIP Codes with some of the higher participation rates are not improving as fast as those with lower initial participation rates. In other words, while the shifting density distribution shows an overall growth, that growth is not observed uniformly across all ZIP Codes. If we look at the last distributional plot, we do see a small bump just below 100 percent. This is quite promising in that is does indicate at least some ZIP Code populations are taking near-full advantage of the Saver's Credit.

From these density distributions, we uncover some important insights into the expansion of participation and the credit's correlation with retirement contributions. In general, as time passes and the eligibility population narrows, we see improvements in credit participation. This suggests there is a time component to the credit's effectiveness. Perhaps more insightful though, is the underlying distribution these figures expose. In each case the distribution has a positive skew, but that skewness diminishes as the eligibility group narrows until Subgroup 2, in which it turns altogether. This means that the national average of tax filers claiming the credit obfuscates the true activity across the U.S. For one, the percent of tax filers per ZIP Code most often observed is slightly smaller than the reported national average but shifts closer to the average over time. For Subgroup 1, while the distribution shifts upward, the shape of the distribution remains relatively constant, this suggests a uniformity in growth overtime. However, for Subgroup 2, not only does the distribution shift upward, but the shape of the distribution narrows. In other words, the variance is narrowing. Where we would expect ZIP Codes in Subgroup 2 to be shifting towards 100 percent, instead higher claiming activity areas remained rather stagnated. While these distributions are insightful, it provides no information into where areas of high and low activity are located.

Section 4.2 - Spatial Distribution

If individuals are static in their understanding of the policy and incentive structure, then any patterns observed spatially should not expand, contract, or shift over time. We have already observed that participation and contribution patterns do shift over time, but we have not actually identified places of growth or stagnation. With that in mind, I turn to thematic maps to pinpoint those geographic areas accounting for the movement in the density distributions, as well as to detect areas of expansion or contraction. For this investigation, I focus on the Southeast region of the U.S., to include Delaware, Maryland, District of Columbia, Virginia, West Virginia, the Carolinas, Florida, Georgia, Kentucky, Tennessee, Mississippi, and Alabama.

Figures 4.3A and B report the quintile distribution of Saver's Credit participation (4.3A) and retirement contributions (4.3B) for the full population in 2002 and 1999 respectively. The first year of the credit's availability saw low levels of take up, with nearly all areas reporting less than 10 percent participation. The areas showing the highest rates of participation are those throughout the Appalachian region and areas just west and throughout most of Kentucky. Prior to 2002, making retirement contributions were generally concentrated in urban areas. Between 1999 and 2002, the region shows a nearly across-the-board increase in retirement contribution activity, barring the border between West Virginia and Kentucky which shows no substantive growth.

Because the Saver's Credit targets low and middle-income households, I isolate the remaining analysis to the two targeted Subgroups. For each Subgroup, I divide the distribution of Saver's Credit participation rates into quintiles for 2002 and hold those quintile cutoffs constant over time. This offers the ability to observe growth in any one ZIP Code relative to its initial rate and witness expansions or contractions across ZIP

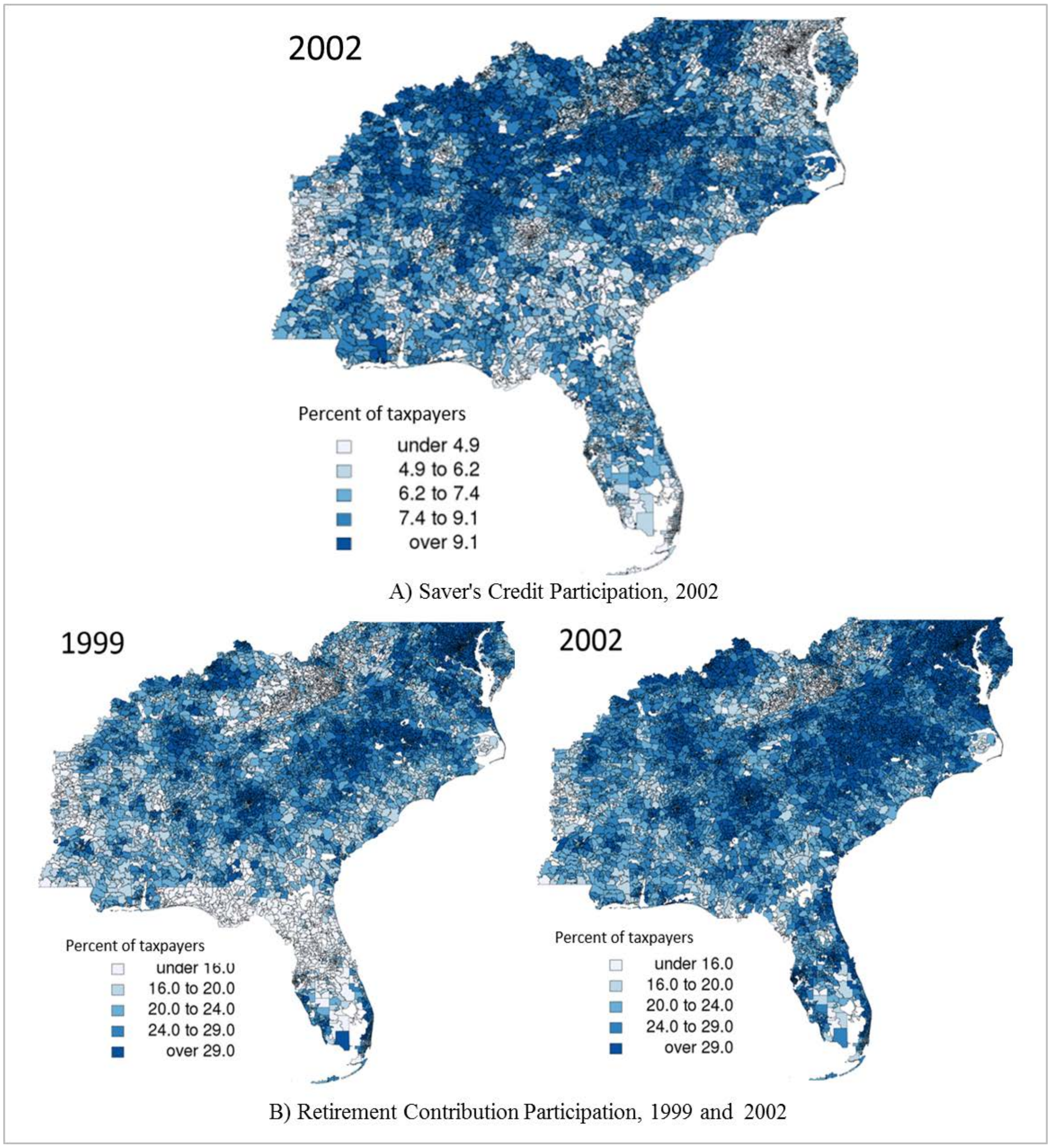


Figure 4.3: Spatial Distribution of Saver's Credit and Retirement Contribution Participation for the Full Population

Codes. Figures 4.4 and 4.5 represent the change in Saver's Credit participation between 2002 and 2013 for Subgroups 1 and 2 respectively.

For Subgroup 1, Figure 4.4 the quintile distribution is heavily concentrated at lower levels, with the top quintile beginning at only 20. I have circled the densest clusters of activity for 2002, which includes the border between Virginia and West Virginia, the borders between Virginia, North Carolina, and Tennessee, and central Kentucky through Tennessee into northern Georgia. Holding quintile cutoffs constant, the expansion of Saver's Credit participation grows rapidly and consistently. This suggests that the credit is growing through either heightened awareness or through the desired effect of inducing more contributions. In either case, we are observing rising participation rates across the board. Even more striking is the pattern of dispersion; improvements in participation appear to flow from prior year density clusters. This has important implications that I discuss further in Section 4.4 and directly measure in Chapters 5 and 6.

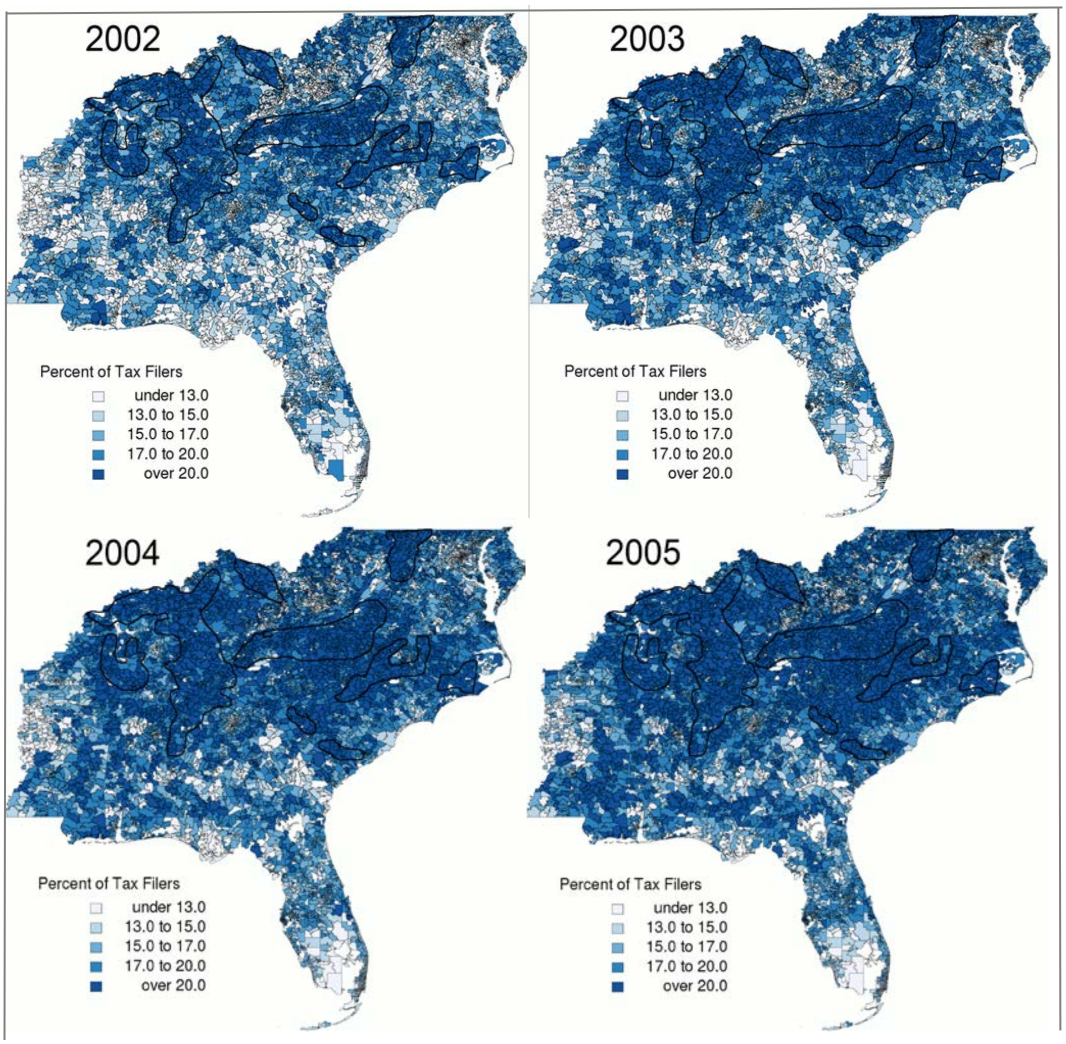
Subgroup 2 is presented in Figure 4.5. While we ought to see 100 percent participation in each year, we already know this is not the case. Relative to Subgroup 1, the sizes of the initial density clusters are smaller but still generally concentrated in the same regions of the map, with two discernable exceptions both within Florida. There are higher concentrations in both central and south Florida. By 2006 the distribution largely shifts to reporting most areas with 60 percent of tax filers or more claiming the credit. The slowest growth appears in a lateral band encompassing southern Alabama and Georgia, and northern Florida. Because the population underlying the figure was restricted to only filers making contributions to a retirement plan, the expansion in credit participation can only imply growth in credit awareness. To extrapolate the velocity of awareness expansion, we need only look at the change between and across years. By 2004 nearly all ZIP Codes have reported at least 44 percent participation; by 2008, nearly all ZIP Codes have reported at least 55 percent participation; and by 2013, nearly all report at least 60 percent. This indicates that knowledge diffusion surrounding the credit was rather slow, having not even fully saturated all areas after more than 10 years.

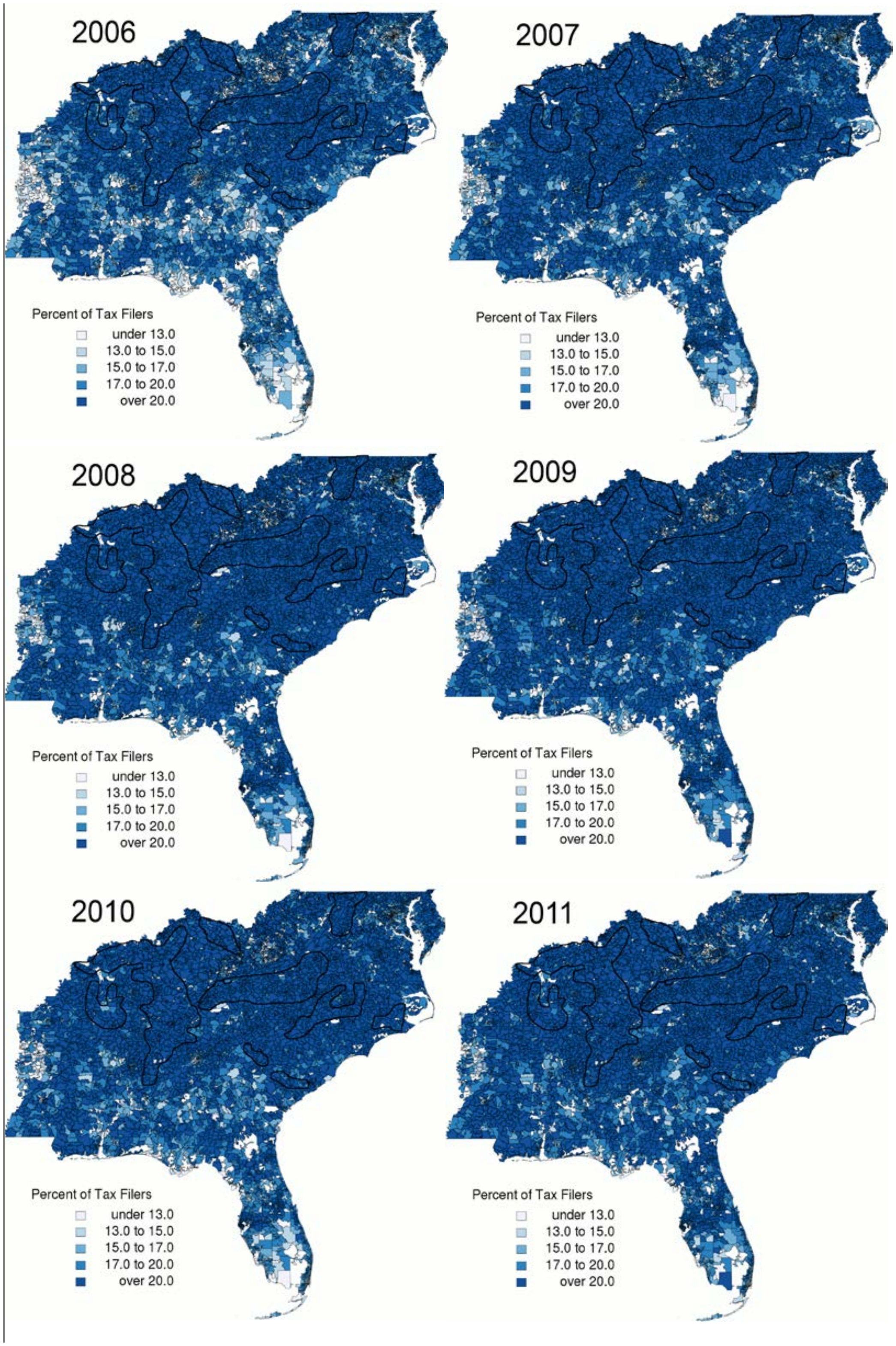
These findings support those reported earlier in Figure 4.1. Though, this adds additional depth to our understanding of the growth of the Saver's Credit over time. We now know that participation was initially concentrated along the southern border between Virginia, West Virginia, and North Carolina, among central Kentucky, Tennessee into northern Georgia, and generally radiated out from these clusters to encompass nearly all areas by 2013.

While these results illuminate the precise areas of higher or lower activity, the purpose of the credit was to improve retirement savings contributions. To explore the changes in contributions, I present Figure 4.6 that reports the percent of the ZIP Code population making a retirement savings contribution for Subgroups 1. Here, I also include two years prior to the introduction of the credit to highlight pre-credit trends.

Figure 4.6 explores the percent of tax filers making a retirement contribution for Subgroup 1. I again fix the distribution to quintiles reported in the base year, here 1999, and encircle the areas with the most density. We see an across the board drop between 1999 and 2000 but large recovery and some expansion by 2002. Between 2002 and 2004 there was very little shift in activity. Between 2004 and 2005 we begin to see some expansion and a new cluster emerging in eastern Mississippi. It is not until 2008 that retirement contribution activity really expands. However, the expansion is short lived with a small dip reoccurring in 2010 followed by slow growth through 2013. By 2013 there are a few additional clustered areas within eastern Mississippi, southern Georgia, and eastern Virginia. In general, though most areas do report growth over this time frame but not as dramatic as that seen for credit participation.

In general, we do find an expansion of retirement contribution activity from pre-credit years. While the subgroup shows a small contraction just after 2002, there are noticeable expansions from pre-credit trends. Specifically, Subgroup 1 shows growth in eastern Maryland and Alabama, central Florida, and throughout South Carolina and Tennessee. contributions.




2006
Percent of Tax Filers
under 13.0
13.0 to 15.0
15.0 to 17.0
17.0 to 20.0
over 20.0
2007
Percent of Tax Filers
under 13.0
13.0 to 15.0
15.0 to 17.0
17.0 to 20.0
over 20.0
2008
Percent of Tax Filers
under 13.0
13.0 to 15.0
15.0 to 17.0
17.0 to 20.0
over 20.0
2009
Percent of Tax Filers
under 13.0
13.0 to 15.0
15.0 to 17.0
17.0 to 20.0
over 20.0
2010
Percent of Tax Filers
under 13.0
13.0 to 15.0
15.0 to 17.0
17.0 to 20.0
over 20.0
2011
Percent of Tax Filers
under 13.0
13.0 to 15.0
15.0 to 17.0
17.0 to 20.0
over 20.0

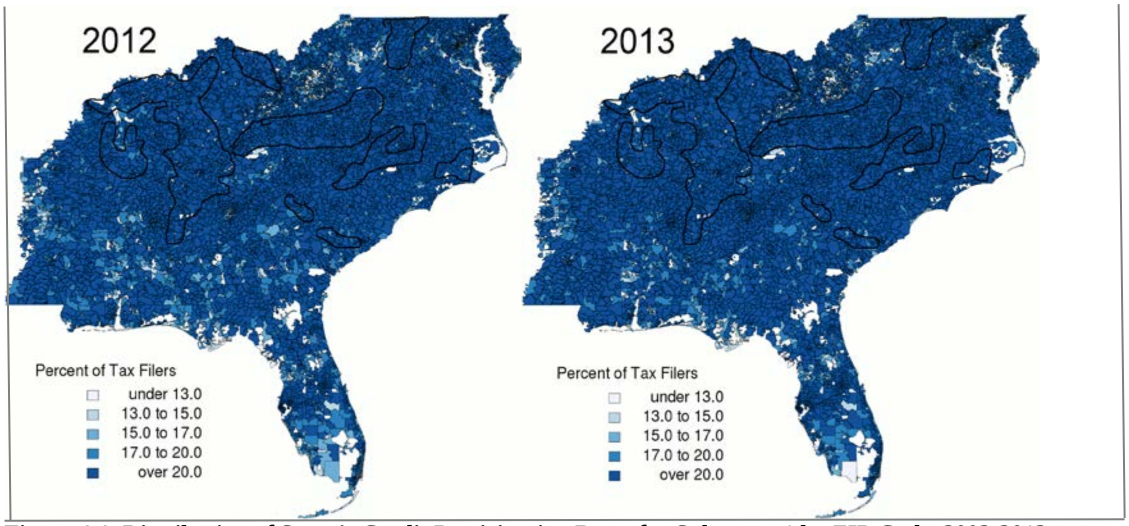
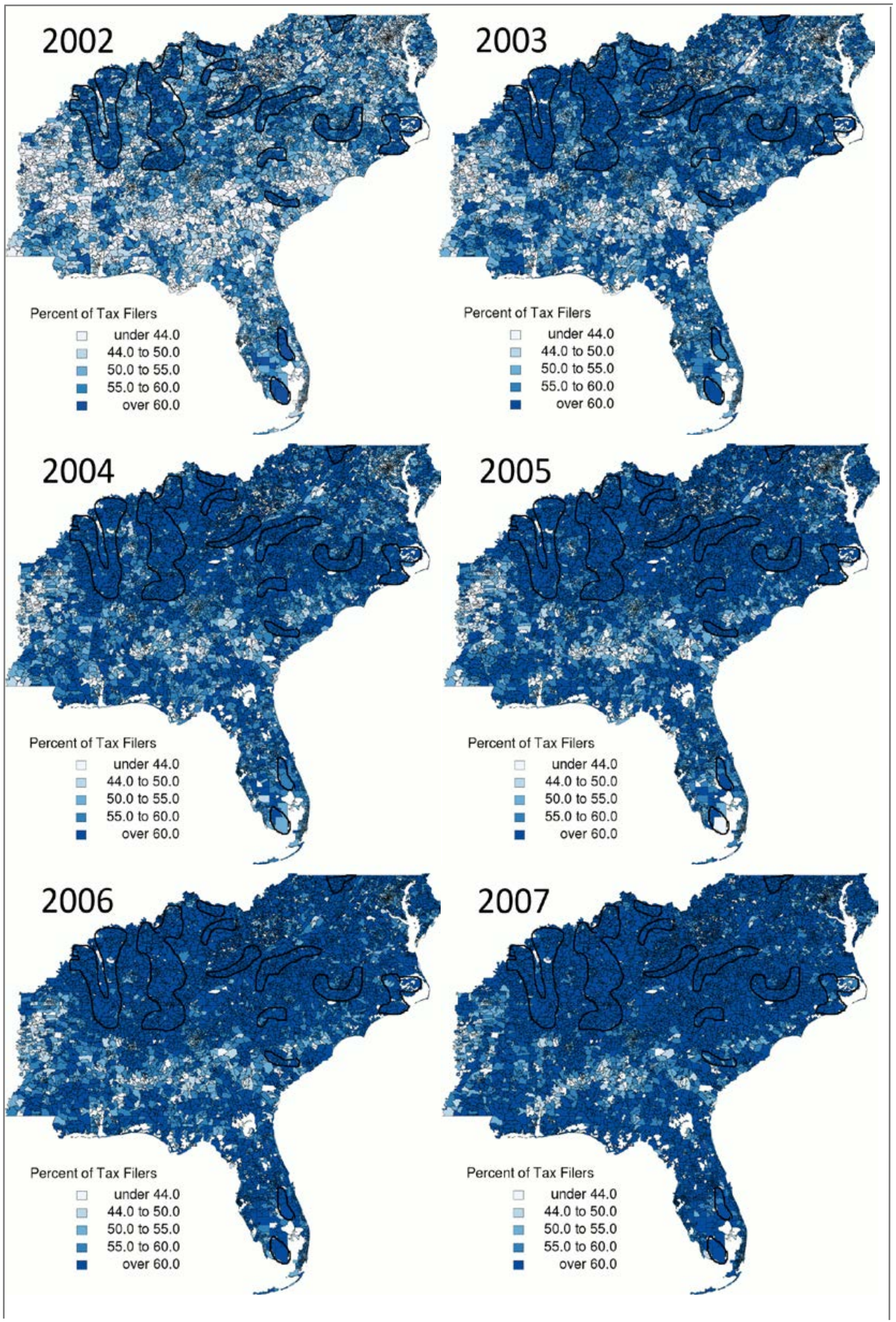


Figure 4.4: Distribution of Saver's Credit Participation Rates for Subgroup 1 by ZIP Code, 2002-2013


2002
Percent of Tax Filers
under 44.0
44.0 to 50.0
50.0 to 55.0
55.0 to 60.0
over 60.0
2003
Percent of Tax Filers
under 44.0
44.0 to 50.0
50.0 to 55.0
55.0 to 60.0
over 60.0
2004
Percent of Tax Filers
under 44.0
44.0 to 50.0
50.0 to 55.0
55.0 to 60.0
over 60.0
2005
Percent of Tax Filers
under 44.0
44.0 to 50.0
50.0 to 55.0
55.0 to 60.0
over 60.0
2006
Percent of Tax Filers
under 44.0
44.0 to 50.0
50.0 to 55.0
55.0 to 60.0
over 60.0
2007
Percent of Tax Filers
under 44.0
44.0 to 50.0
50.0 to 55.0
55.0 to 60.0
over 60.0

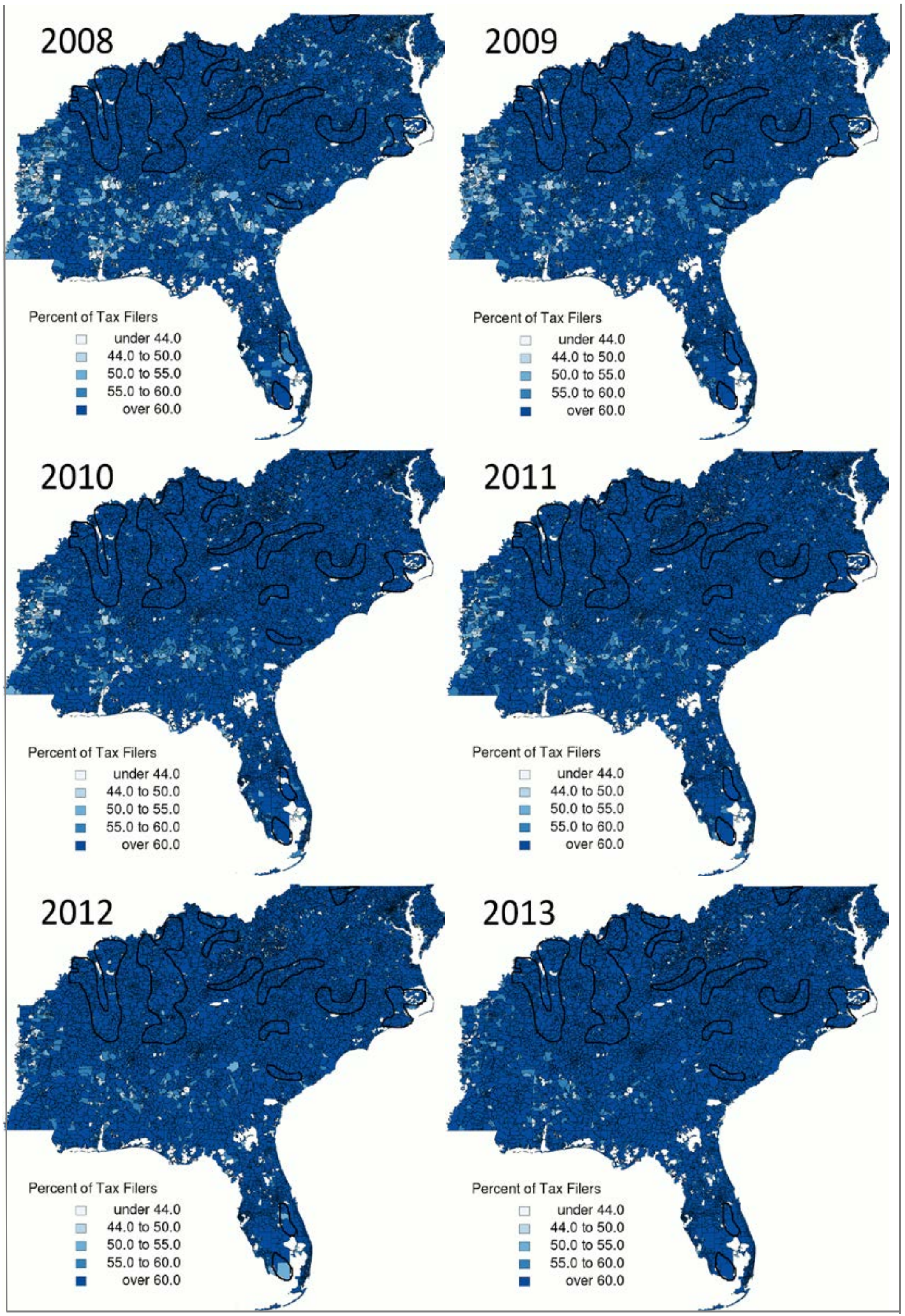
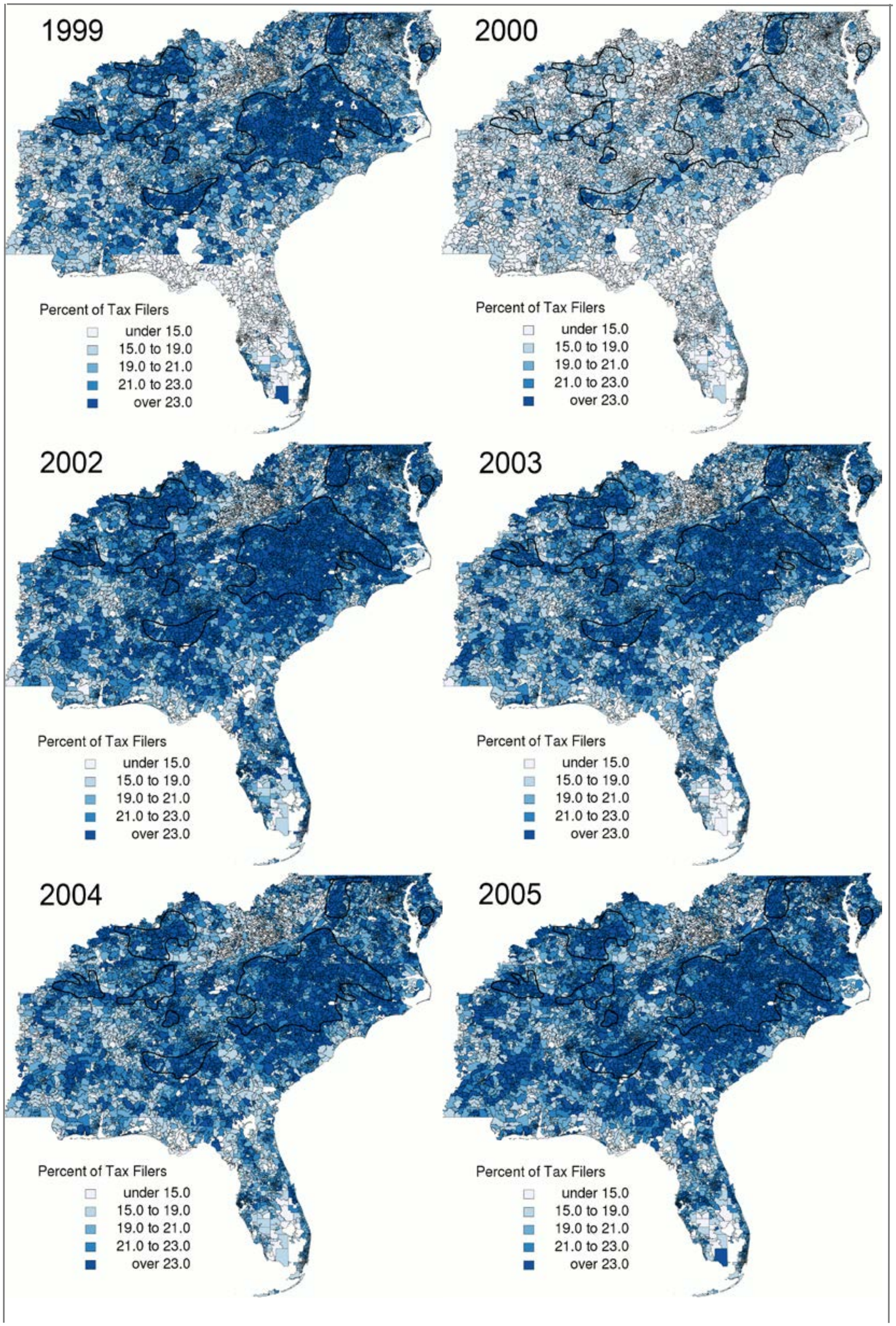
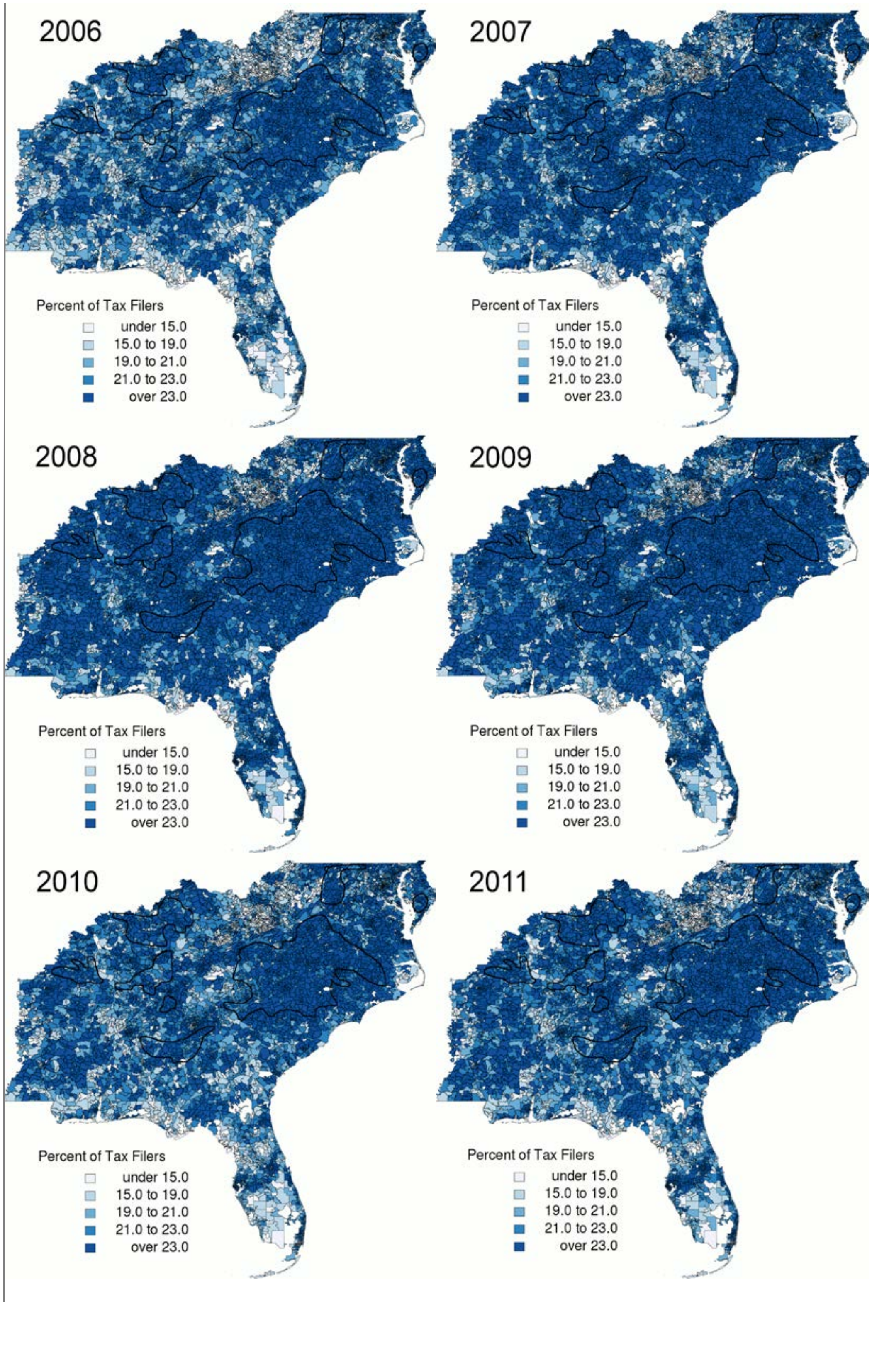


Figure 4.5: Distribution of Saver's Credit Participation Rates for Subgroup 2 by ZIP Code, 2002-2013


1999
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0
2000
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0
2002
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0
2003
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0
2004
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0
2005
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0


2006
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0
2007
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0
2008
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0
2009
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0
2010
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0
2011
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0

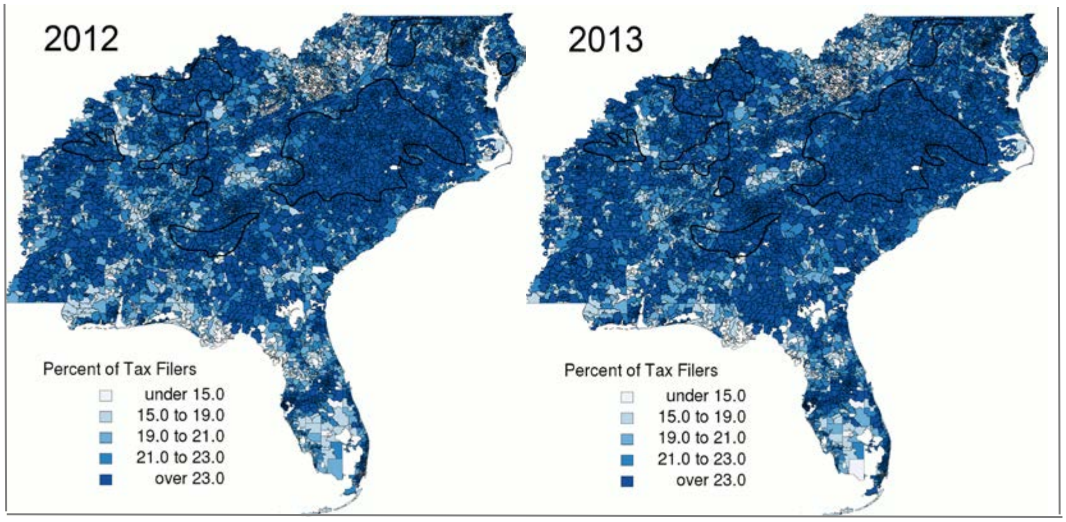


Figure 4.6: Distribution of Retirement Contribution Activity Rates for Subgroup 1 by ZIP Code, 2002-2013

Section 4.3 - Spatial Overlap

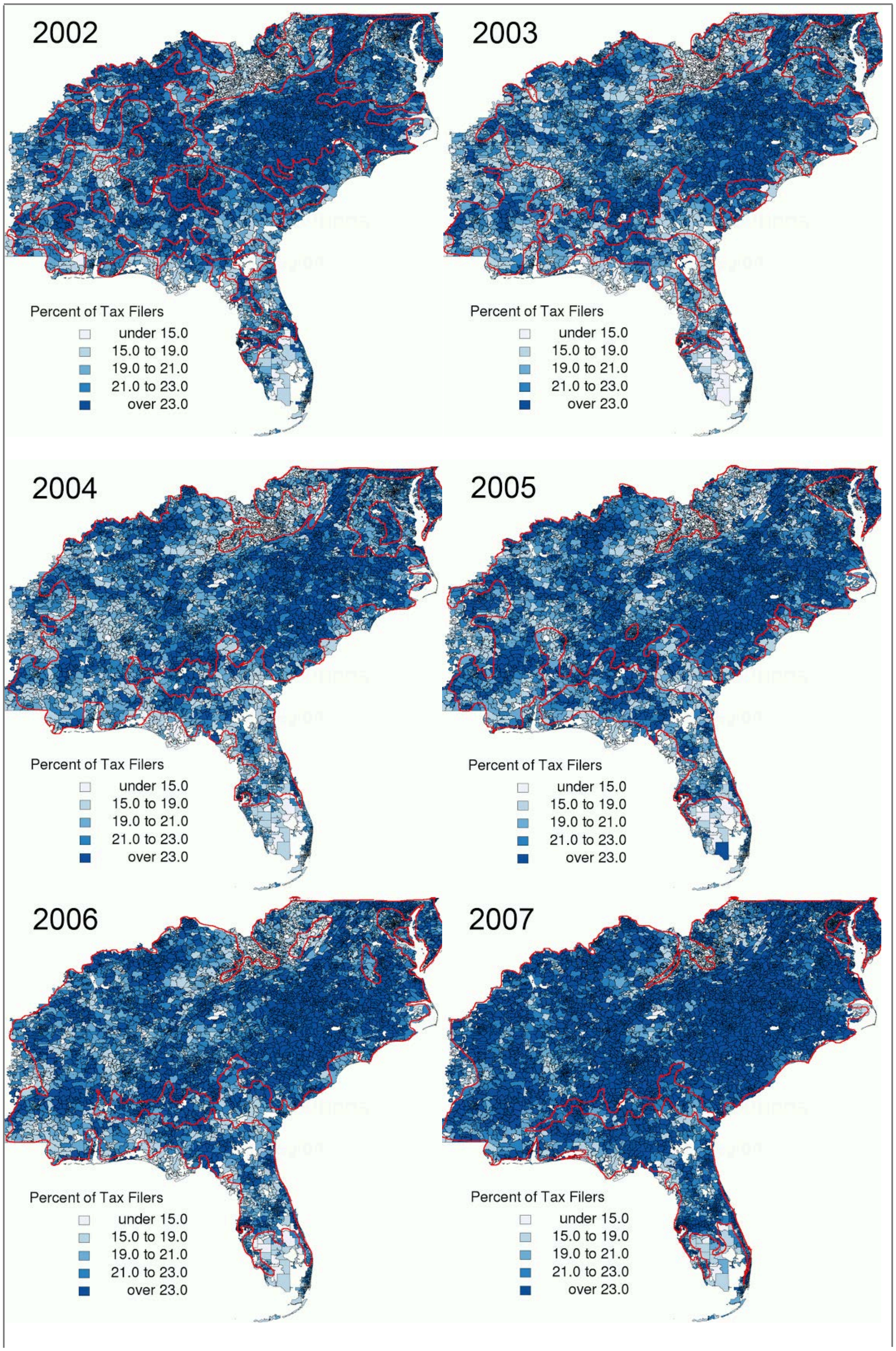
The purpose of the Saver's Credit is to incentivize retirement savings among low to middle-income households. Although we have seen growth in both credit participation and retirement contribution activity over time, thus far we have not explored the relationship between the two. If the Saver's Credit is inducing individuals to make retirement contributions, there must be a positive correlation between the two a priori. To explore this, I overlay the expansion of the Saver's Credit, highlighted in red, onto the map of retirement contribution activity, presented in Figure 4.7.

For Subgroup 1, overall there is substantial overlap overtime between retirement contribution activity and Saver's Credit participation. There are a few inconsistencies, however. For a few areas retirement contribution participation is initially heavily concentrated without overlap in credit participation, particularly areas in eastern North

Carolina, central South Carolina and Georgia, and throughout Florida. Conversely, in 2002 the southern tip of Mississippi and north-west Alabama show higher concentrations of credit participation without correspondingly high contribution activity. Central West Virginia joins this group in 2005 and 2013 along with eastern Kentucky in 2009 and 2010. Areas having high retirement contribution activity not directly aligned with credit participation could still be the result of credit availability. Such is the case for tax filers who respond to the incentives of the credit but discover they have no tax liability upon completing their tax return for the year. Areas indicating high credit participation without correspondingly high rates of retirement contributions on the surface appear to violate the necessary condition for establishing causality between the credit and retirement contributions.

On the surface these disparate areas might be cause for concern. However, one must keep in mind how these statistics were derived. Recall, Subgroup 1 is all tax filers that had income below the Saver's Credit income threshold and with tax liability. When computing the distribution of tax filers claiming the credit and the distribution contributing to a retirement plan the underlying population totals do not change between the two analyses. In fact, recall that in order to claim the credit one must have contributed to a retirement plan, thus making credit participation in effect a subset of the population making a retirement contribution. Said differently, the denominator used for computing the percent of credit participation and retirement contribution activity do not change, and the numerator for computing the percent of credit participation is necessarily less than or equal to the numerator for percent of retirement contribution activity. Therefore, those areas for which Saver's Credit participation is observed to be higher than those for retirement contribution activity is be due to the slight differences in quintile distribution between the two but exaggerated thematically.

In general, the overlap between Saver's Credit participation and retirement contribution activity is quite strong for Subgroup 1. This indicates a quite high positive correlation between the two. And while not an indication of causality per se, such correlation is a necessary condition in the process towards establishing causation. To be certain, these graphs provide no information on relationship directionality. It is unclear whether the correlation is driven by incentives induced through the Saver's Credit leading to growth in retirement contribution activity or if growth in activity results in more tax filers claiming the credit after the fact. I discuss this direction further in Chapter 7.


2002
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0
2003
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0
2004
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0
2005
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0
2006
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0
2007
Percent of Tax Filers
under 15.0
15.0 to 19.0
19.0 to 21.0
21.0 to 23.0
over 23.0

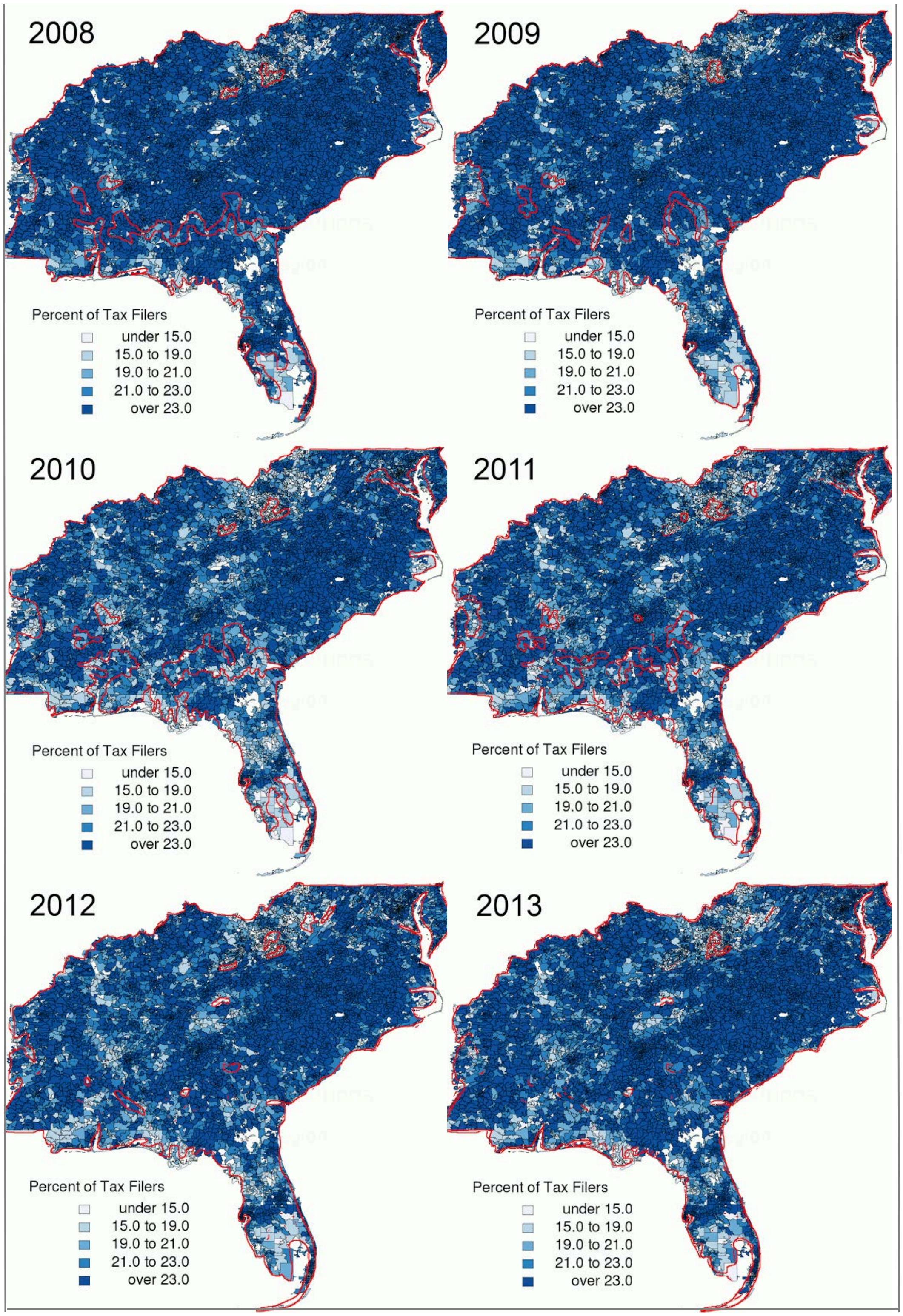


Figure 4.7: Overlap between Saver's Credit and Retirement Contribution Participation for Subgroup 1 by ZIP Code, 1999-2013

Section 4.4 - Discussion

Earlier research exploring the utilization and effects of the Saver's Credit reported a lackluster response, but each suffered from restricting analysis to a short time horizon and having only national representation. Here, I go beyond the previous work to present new evidence that documents growth over a longer time horizon, locate pockets of high participation, and map patterns of dispersion. Through this, I reveal an overall improvement in participation of filers taking advantage of the Saver's Credit both over time and across populations of eligibility. For the full population, we do see positive shifts in participation over time but the range across locality diminishes. For more targeted populations, we see a far more uniformly positive shifting over time and indications that the pace of expansion is well above what would be expected when relying on the full population alone. For those filers meeting all eligibility requirements for claiming the credit, Subgroup 2, we observe larger growth over time, a larger area of concentrated participation initially, and expansive utilization in the final few years. These results offer up new insights into the reception of the Saver's Credit. What we observe is an ever-increasing awareness and utilization of the credit that shorter time horizon analyses cannot capture. Where previous research may have disregarded the benefits of the credit initially, this research demands a new investigation into their conclusions now that a larger proportion of the population has become aware.

Beyond just capturing the distributional shifts in participation, this study identifies locations of high and low participation throughout the Southeast region of the U.S. and across the various targeted populations. Initially, the Saver's Credit enjoyed high utilization concentrated in a "V" shape along the southern borders between Virginia, West Virginia, and North Carolina, Tennessee through central Kentucky and in northern parts of Georgia. As I restrict the population of eligibility, this generality persists but expands to include pockets of Florida and parts of southern Mississippi, Alabama, and Georgia. Areas indicating particularly low participation include western Mississippi, central Alabama and Georgia, and southern West Virginia. This pattern holds for all the definitions of eligibility. Knowing this helps to identify precise areas for protentional policy outreach and advocacy, or perhaps further policy intervention.

Even more, I present the distribution of retirement contribution participation for these same areas. This further highlights locations for possible outreach. Because the credit's goal is to encourage retirement savings, identifying areas reporting low retirement contribution activity pinpoints locations that would most benefit from a targeted outreach approach. Areas with some of the lowest participation include most parts of West Virginia, western Mississippi, and southern Georgia.

In addition to revealing initial states of behavior, I also document patterns of dispersion over a long horizon. Where initially Saver's Credit participation was generally isolated to the "V" shape pattern discussed above, as time progressed, we observe expansion that seems to radiate out from those initial density clusters. This holds true for each of the eligibility groups. Even for those areas that were not identified in the initial density cluster, we see this radiating effect taking place. As one location increases participation, surrounding areas appear to follow suit with an observable lag. This is more so the case for Saver's Credit participation than retirement contribution participation. And as such, this pattern of behavior speaks to where targeted outreach would be most beneficial. Any outreach program would be best served to locate in areas further from these density clusters to allow for maximum penetration. By knowing that areas surrounding already dense activity will in short order begin showing similar characteristics, focusing outreach on areas further away inject new information in areas least likely to obtain it on their own. I find that areas with consistently low participation include southern West Virginia, western Mississippi, and a lateral band encompassing southern Alabama and Georgia.

While this radiating effect appears for only credit participation, it does show strong evidence of a positive correlation between credit and contribution participation. Although it is clear there is a positive relationship between the two, it is unclear in which direction the impact flows. It may be the case that the Saver's Credit is improving contribution participation, or it could be individuals are responding to other incentives that enhance contribution participation and only claiming the credit after they had already chosen to contribute, or there may still be a confounding factor driving the observance of both behaviors. I explore this further in Chapter 7.

In the next Chapter I turn to explicitly measuring the existence and magnitude of spatial relationships that up to this point I only infer.

CHAPTER 5 - THE SPATIAL DIMENSION OF PARTICIPATION IN THE SAVER'S CREDIT

Where the last study could only suggest a relationship between location and Saver's Credit participation, in this next study I explicitly quantify the concentration and dispersion of spatial correlation among credit participation. Here, I measure the correlation of participation across ZIP Codes to determine if said participation is at least in part dependent on where one resides. Such information is critical if one is to understand the mechanism underpinning the observed expansion of Saver's Credit, or to construct a successful outreach program. To do this, I first compute the Moran's I statistic for the continental U.S. and the Southeast region. Going beyond this global average, I then disaggregate spatial autocorrelation by calculating the local Moran's I statistic at the ZIP Code level for the Southeast region and map these results.

Such explicit estimations of spatial autocorrelation is relatively new to the field of tax and tax expenditures. Only a handful of studies have endeavored to explore the relationship between tax and location, few have utilized this particular method for estimating the spatial autocorrelation underpinning tax policy, and even fewer have intertwined the two[11]. This marks the first study to explicitly quantify the spatial autocorrelation underpinning the Saver's Credit and retirement savings participation.

This study proceeds as follows. In the next section I explain how I determine limits for spatial proximity, or "neighbors". Then I present Moran's I statistics for both

[11] See Alm and Yunus (2009), Albouy (2009), Chetty et al. (2013), Fajgelbaum et al. (2019), and Adhikari et al. (2019).

Saver's Credit participation and retirement participation through observed contributions. I then disaggregate this to the ZIP Code level for the Southeast region to pinpoint locational variation. Finally, I provide a discussion before moving on to Chapter 6.

Section 5.1 - Spatial Weight Matrix

The first critical step in examining spatial autocorrelation is determining the existence and structure of the relationship of influence between neighboring ZIP Codes. To do this, I derive the spatial weight matrix, *W*. The purpose of such a matrix is to explicitly identify and assign the intensity of influence that neighboring ZIP Codes, *j*, have on the ZIP Code of interest, *i*. Each value within the *n x n* matrix, ω_{ij} represents the assumed influence of *j* on *i* and, given that this is assumed a priori, the formulation of the matrix is of critical importance.

There are two practical approaches for defining the existence and relationship between neighboring ZIP Codes. For one, a neighbor could be defined as one that explicitly shares a border, in which case one must determine which borders matter. Or, a neighbor could be defined based on distance away from ZIP Code *i*. It is reasonable to assume that individuals' movements, and therefore circle of influence, are not limited to only crossing between adjacent ZIP Codes and thus warrants this latter, slightly more computationally taxing, relationship structure. For this study I rely on this latter structure.

Specifically, I employ a d-nearest neighbor weighting approach, relying on the Euclidean distant between ZIP Code centroids to measure nearness. This approach is quite similar to that of the k-nearest neighbor method. But rather than defining the number of observations *k* which are closest to *i*, *d* defines the distance from *i*. The

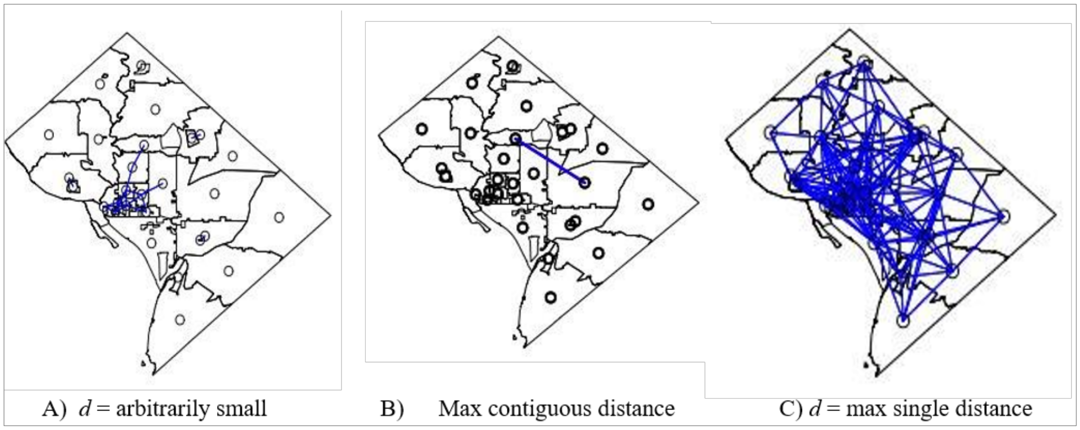


Figure 5.1: Proximity Map for the District of Columbia

advantage is the ability to set a radius of influence. For simplicity, I define *d* as the maximum distance between an adjacent *i*,*j* combination. In other words, *d* is equal the longest distance between any two border-sharing ZIP Codes[12].

To see how this works, I present the proximity map for the District of Columbia (D.C.) in Figure 5.1A-C. The District of Columbia has only 28 ZIP Codes so it offers a convenient region to observe how the spatial weight matrix is constructed. The circles in each figure identify ZIP Code centroids and are the basis for measuring the distance between location. The blue lines connect neighboring centroids and define proximity. Figure 5.1A presents the proximity map for an arbitrarily small *d* distance parameter. Here, only closely neighboring centroids have been identified as connected. Figure 5.1B

[12] The Department of Transportation (DOT) estimates that the average personal trip length is somewhere between 9.39 and 10.11 miles per day and has remained relatively constant over the past 40 years (2011). Of this, the majority is reserved for social and recreational activities, peaking at an average of 12 miles per day in 2009. The maximum distance, *d*, I observe is approximately a 30-miles radius. While longer than the 12-mile radius suggested by (DOT), this affords two advantages. First, the 12-mile radius was an average suggesting that for many, this limit is overly binding, and this radius only accounts for distances traveled. Most individuals are influenced well beyond this radius through media, social media, telecommunication, and other forms of networking outside physical location. Having a distance larger than 12-miles allows for the capturing of some of that non-tangible influencing. Second, it ensures all ZIP Codes have at least one connected neighbor, affording easier computation.

highlights the maximum distance between a border-sharing adjacent *i,j* combination, which defines *d*. Figure 5.1C shows the complete proximity map, with *d* set equal to the distance represented in Figure 5.1B. There are 326 blue lines traversing this small map. Expanding this process to the whole of the Southeast region, 671,240 neighbors were identified as connected. The ZIP Code with the least neighbor connections was 1, by definition, and the most had 7,496, with a mean of 3,969 and median of 4,281 connections. The full continental U.S. had 25,311,968 neighbor connections, with a minimum of 1, maximum of 31,549, mean of 14,713, and median of 14,694.

Once proximity has been identified, the weight matrix can be populated. The exact ω_{ij} element assignment can be either binary or inversely related to distance. In the binary case, 1 identifies neighbors and 0 otherwise. In the inverse-distance method, the longer the distance, the smaller the weight assignment will be, ranging from 1 to 0. For this research I explored the results using both proximity matrices and found little to no difference in outcome but significant difference in computational efficiency. Thus, for the remainder of my research I utilize binary weighting assignments. This exercise was conducted for each tax year of analysis.

Section 5.2 - Global Spatial Autocorrelation

Now that the spatial weight matrix has been populated, I can begin analysis by assessing the global spatial autocorrelation using the Moran's I statistic. This statistic is in effect a correlation coefficient that estimates the multidimensional autocorrelation between location *i* and *j* for all neighbor combinations. The outcome is a statistic ranging between -1 and 1 that averages the measured covariance across *ij* combinations. An estimate of -1 implies perfectly negative correlation, or areas that repel similar activity in neighboring areas. An estimate of 1 implies perfectly positive correlation, or areas that perfectly attract similar activity in neighboring areas. This is all relative to what would be expected through random chance. Formally,

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{ij}(z_i - \bar{z})\ (z_j - \bar{z})}{S_z^2\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}}$$

where z is the variable of interest, and S_z^2 is the variance. Simply put, the Moran's I measures the covariance between *i* and *j* divided by the sample variance. Intuitively, if spatial autocorrelation were ignored, thus omitting ω_{ij} from the above equation, we would ostensibly be left with the Pearson correlation coefficient.

I ran this for the percent of filers claiming the Saver's Credit and the percent of filers making a retirement contribution for both the U.S. and the Southeast Region, with results shown in Table 5.1. Two interesting patterns emerge. First, every statistic indicates spatial autocorrelation regardless of year, eligibility definition, or variable of interest, all at or above the 0.05 statistical significance level[13]. But as we move toward more targeted eligibility groups, the intensity diminishes. Averaging across all years, the U.S. as a whole reports a Saver's Credit Moran's I of 0.3429 for the full population, 0.2703 for Subgroup 1, and 0.0843 for Subgroup 2. This pattern of progressively smaller spatial autocorrelation holds for retirement contribution participation as well, averaging 0.29589 and 0.23518 respectively. These findings are quite illuminating. This shows that the concentration of participation is progressively less acute the more stringent our

[13] Recall, this statistic is computed relative to the expected correlation coefficient of zero, thus the null hypothesis is $H_0 = 0$.

Table 5.1 Moran's I Statistic by Eligibility Group and Geographic Region, 2002-2013[1]

	Saver's Credit Participation			Retirement Contribution	
Year	Population	Subgroup 1	Subgroup 2	Population	Subgroup 1
Continental U.S.					
Average	**0.3429**	**0.2703**	**0.0843**	**0.29589**	**0.23518**
2002	0.3752	0.3253	0.1656	0.30238	0.24178
2003	0.3720	0.3143	0.1552	0.31352	0.24342
2004	0.3559	0.2965	0.1417	0.31547	0.23183
2005	0.3539	0.2913	0.1218	0.31129	0.23700
2006	0.3477	0.2793	0.0985	0.31162	0.23744
2007	0.3449	0.2743	0.0773	0.31821	0.23922
2008	0.3506	0.2702	0.0638	0.28869	0.24103
2009	0.3505	0.2658	0.0505	0.28583	0.25603
2010	0.3383	0.2539	0.0497	0.28403	0.23382
2011	0.3194	0.2352	0.0398	0.27656	0.22393
2012	0.3043	0.2229	0.0267	0.27305	0.22219
2013	0.3027	0.2147	0.0210	0.26996	0.21449
Southeast Region [2]					
Average	**0.4816**	**0.2604**	**0.1367**	**0.62154**	**0.28559**
2002	0.4980	0.3090	0.1539	0.69961	0.31316
2003	0.4871	0.2953	0.1458	0.72996	0.32614
2004	0.4990	0.3044	0.1932	0.72476	0.29978
2005	0.4933	0.2874	0.1973	0.68838	0.29479
2006	0.4948	0.2699	0.1589	0.66423	0.28237
2007	0.4841	0.2501	0.1368	0.72198	0.26767
2008	0.5261	0.2810	0.1429	0.57630	0.28690
2009	0.4966	0.2689	0.1236	0.55351	0.27463
2010	0.4856	0.2465	0.1270	0.54882	0.25761
2011	0.4535	0.2238	0.1088	0.53412	0.26775
2012	0.4375	0.2008	0.0827	0.49639	0.28177
2013	0.4231	0.1879	0.0688	0.52041	0.27447

[1] All statistics are significant at the 0.001 level
[2] Southeast region includes Delaware, Maryland, District of Columbia, Virginia, West Virginia, North and South Carolina, Florida, Georgia, Kentucky, Tennessee, Mississippi, and Alabama
Source: author's computation

definition of eligible becomes. So, while the credit may not be heavily utilized nationally, the awareness of the credit is not uniformly distributed.

The estimates for the Southeast region are consistently larger than the U.S. The explanation for this is not clear. It could be the case that geographic size of the underlying ZIP Codes could play a role in this, as smaller land masses are more likely to have a higher number of connected neighbors. It could also be the case that information is more easily transferred across adjacent borders. Or it may be the case that there is an unobservable social network effect that resides in this region of the U.S. but not others. It is unclear exactly why this discrepancy exists and begs further investigation.

Putting this aside, Table 5.1 also highlights an interesting temporal pattern. For any given eligibility group, spatial autocorrelation diminishes over time. For Saver's Credit participation, spatial correlation is estimated at 0.3752 for the full population in 2002, by 2013 this had fallen to 0.3027, representing a nearly 20 percent decline overtime.

These trends are more easily seen through Figures 5.2A and B plotting the Moran's I for the continental U.S. and Southeast region respectively. The solid lines track Saver's Credit Moran's I estimates and the dotted lines track retirement contribution participation estimates. Each eligibility group is assigned a different color with estimates for the full population presented in blue, estimates for Subgroup 1 in green, and Subgroup 2 in red. Of note, and by design, Subgroup 2 shows no estimates for participation in a retirement plan. For the U.S., Figure 5.2A, spatial autocorrelation is consistently estimated higher for Saver's Credit participation than for retirement contribution

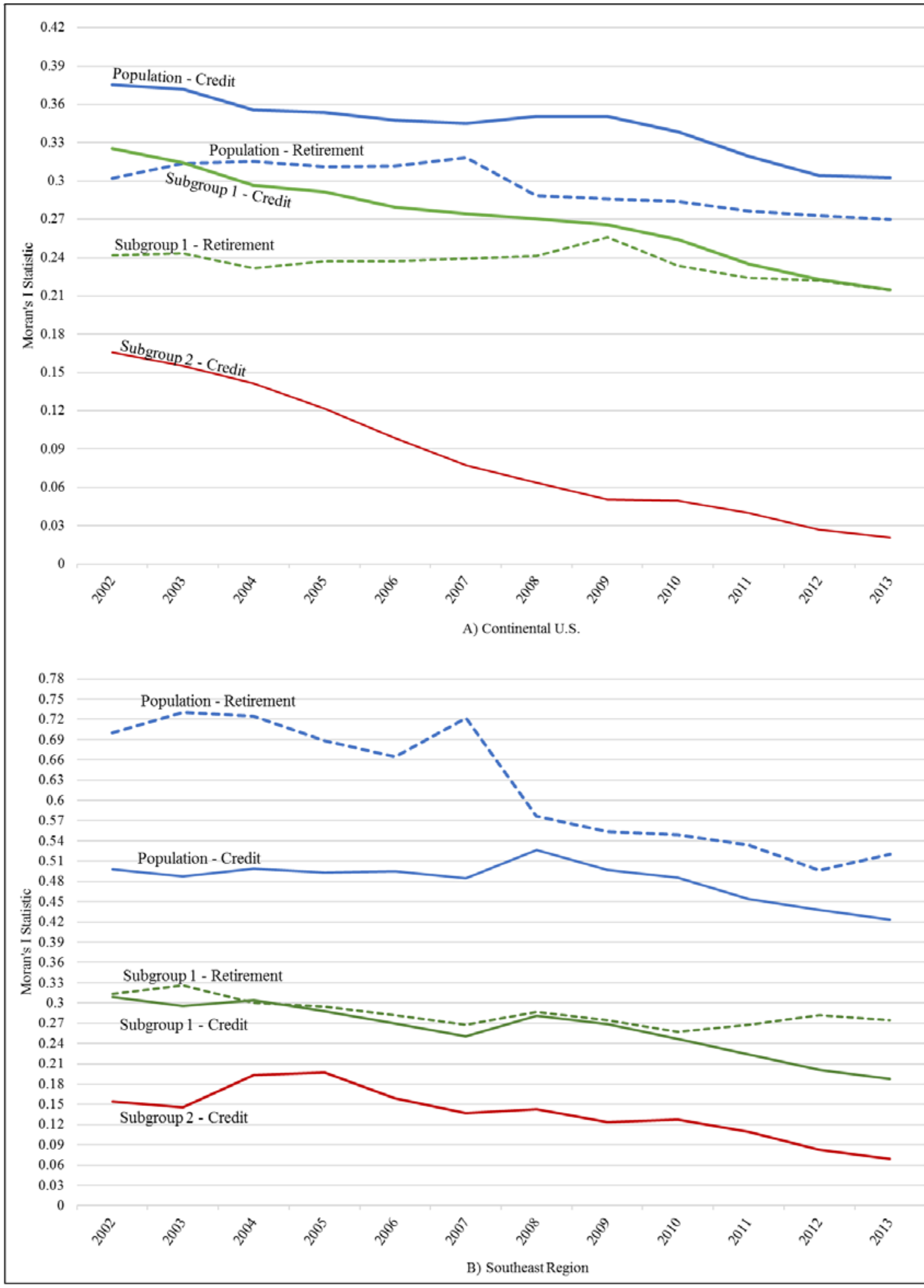


Figure 5.2: Moran's I Statistic for Saver's Credit and Retirement Contribution Participation by Eligibility Group and Geographic Region, 2002-2013

participation, though the gap lessens over time. Turning to Subgroups 1 and 2, the rate of decline for the Saver's Credit estimates is not uniform, as the population becomes more targeted the rate of decline increases. This suggests that rate of knowledge dispersion grows at an ever-increasing rate the more targeted the population of interest becomes. These patterns are not observed for retirement contribution participation. Here instead, we see generally no substantive declines until 2008, for the population when it drops from 0.318 to 0.289, and in 2010 for Subgroup 1, when it drops from 0.256 to 0.234.

Switching to the Southeast region, Figure 5.2B, we observe different patterns than those of the U.S. While we again see a decline over time and across eligibility groups, retirement contribution participation estimates are consistently at or above those for Saver's Credit, with only a few minor exceptions. In general, the rate of decline for Saver's Credit participation is more uniform across the groups than the larger U.S. But unlike the U.S., the Saver's Credit appears to have upticks in 2004 and 2008. Subgroup 2 shows the largest jump for 2004, going from 0.1458 to 0.1931, whereas Subgroup 1 has the largest for 2008, from 0.2501 to 0.2810. It is unclear what is causing these irregular patterns.

Turning to the marginal rate of change, Table 5.2 evaluates the rate of change in the Moran's I between years to help infer a velocity, or at least volatility, of dispersion. The average marginal decline of the correlation coefficient for the full population is approximately 0.0066 for the U.S, and 0.0068 for the Southeast region with respect to Saver's Credit participation. More interesting though, is the trend over time. The average decline conceals more volatility between years. While in general there is a trending

Table 5.2 Moran's I Statistic Marginal Change by Eligibility Group and Geographic Region

	Saver's Credit Participation			Retirement Contribution	
Year	Population	Subgroup 1	Subgroup 2	Population	Subgroup 1
Continental U.S.					
Average	**-0.0066**	**-0.0101**	**-0.0132**	**-0.0029**	**-0.0025**
2003	-0.0032	-0.0110	-0.0104	0.0111	0.0016
2004	-0.0161	-0.0177	-0.0135	0.0019	-0.0116
2005	-0.0020	-0.0052	-0.0198	-0.0042	0.0052
2006	-0.0062	-0.0120	-0.0233	0.0003	0.0004
2007	-0.0028	-0.0050	-0.0212	0.0066	0.0018
2008	0.0057	-0.0041	-0.0135	-0.0295	0.0018
2009	-0.0002	-0.0043	-0.0133	-0.0029	0.0150
2010	-0.0121	-0.0119	-0.0008	-0.0018	-0.0222
2011	-0.0190	-0.0188	-0.0099	-0.0075	-0.0099
2012	-0.0151	-0.0123	-0.0130	-0.0035	-0.0017
2013	-0.0016	-0.0082	-0.0058	-0.0031	-0.0077
Southeast Region					
Average	**-0.0068**	**-0.0110**	**-0.0077**	**-0.0163**	**-0.0035**
2003	-0.0109	-0.0137	-0.0081	0.0303	0.0130
2004	0.0119	0.0091	0.0474	-0.0052	-0.0264
2005	-0.0057	-0.0171	0.0041	-0.0364	-0.0050
2006	0.0015	-0.0175	-0.0384	-0.0242	-0.0124
2007	-0.0107	-0.0197	-0.0222	0.0577	-0.0147
2008	0.0420	0.0308	0.0061	-0.1457	0.0192
2009	-0.0295	-0.0121	-0.0193	-0.0228	-0.0123
2010	-0.0110	-0.0224	0.0034	-0.0047	-0.0170
2011	-0.0321	-0.0227	-0.0182	-0.0147	0.0101
2012	-0.0160	-0.0230	-0.0261	-0.0377	0.0140
2013	-0.0144	-0.0129	-0.0139	0.0240	-0.0073

Source: author's computation

negative behavior to the measured correlation coefficient, the decline is not uniform for either variable of interest or geographic region. Focusing strictly on the Saver's Credit, Subgroup 1 appears to follow a similar trajectory over time with volatility a bit more amplified for the former. Subgroup 2 shows a far more consistently growing decline in change until a shift between 2009 and 2010. While still dropping, the year-over-year change is much smaller than any year prior or post. The Southeast region shows even more volatility, for all eligibility definitions. While true, the volatility pattern across eligibility groups appears generally consistent.

The results from this global statistic confirms the implications drawn from my previous study, that participation in the Saver's Credit is not spatially independent. This indicates that values observed in one ZIP Code are systemically related to values observed in neighboring ZIP Codes, on average. These results are robust to other measures of spatial association[14]. While this research explores only the Saver's Credit, a generally touted inconsequential tax expenditure, given the preponderance of evidence that spatial autocorrelation exists, it is reasonable to assume that other credits will show a similar pattern of acute initial spatial dependence with diffusion over time.

There is an important aspect that cannot be overlooked, the Moran's I explores the spatial autocorrelation by averaging across the whole of the geographic area. That is, it conceals the varying distributional patterns of participation across ZIP Codes that drive these results. Moreover, it cannot identify exact locations that drive the statistic in one direction or another. I turn to expounding on locational clustering in the next section.

Section 5.3 - Local Indicators of Spatial Association, LISA

To assess the local patterns underpinning the global spatial correlation statistics, one must disaggregate those statistics down to the local level by way of Local Indicators of Spatial Association (LISA) statistics. LISA statistics afford the opportunity to decompose the global static into its varying locations with the attractive feature of being proportional to the global statistic (Anselin, 1995). Among the leading methods applied

[14] See Appendix A for further discussion.

in the literature today is the local Moran's I[15]. The local Moran's I, by definition, is quite similar to its global counterpart. Instead of summing across locations *i* and *j*, the local Moran's I produces an estimate for each *i*. This allows for the identification of specific locations that express higher/lower concentrations than would be expected by random chance. Formally, it is defined as

$$I_i = \frac{\sum_{j=1}^{n} \omega_{ij}(z_i - \bar{z})\,(z_j - \bar{z})}{S_z^2 \sum_{j=1}^{n} w_{ij}}$$

Notice, the only difference from that of the Moran's I is the omission of the summation across *i*. The interpretation is the same as before, with values closer to 1 indicating more clustering and values closer to -1 indicating more dispersion, relative to what would be expected randomly. Now each ZIP Code will have an associated estimate of correlation to their connected neighbors.

Because we now obtain one estimate for every ZIP Code, it is unrealistic to present every estimate produced. Instead, I map the local Moran's I estimates for the Southeast region exclusively. I look at high and low participation in Saver's Credit clusters between 2002 and 2013 and across eligibility groups, Figures 5.3 A-C. Appendix B analyzes the robustness of these results.

Figure 5.3A presents the cluster map for the full population. Each area highlighted in either red or blue signifies a statistically significant local positive correlation to a

[15]Other measures include the Getis-Ord G_i^* and local Geary G_i^* measures the deviation of location *i* from that of the global average. The local Geary decomposes the global estimate similarly to the local Moran's I. For this research I focus on the local Moran, as it is not clear that the G_i^* would lead to unbiased inferences given the significant variation in the underlying distributions leading to statistically insignificant G* estimates, and the local Geary mirrors the local Moran in decomposition and does not appear to add any additional information.

neighboring ZIP Code at the 0.05 significance level. More precisely, each area in red identifies a ZIP Code that has a highly positive correlation and a high Saver's Credit participation rate, relative to what would be expected by random chance. Conversely, blue areas identify locations that have high correlation but low participation rates. What we observe initially is high participation clusters closely focused around the Appalachian region into and throughout Kentucky and Tennessee. Low participation clusters are generally focused around highly urban areas, such as D.C., Richmond, Charleston West Virginia, Atlanta, and Miami, among others. For the center of the map, the clustering pattern falls generally along the urban/rural divide with urban areas showing low participation clusters and rural reporting high.

As time passes, we observe minimal movement in these clustered areas. There are however, three areas that show interesting changes. There is substantial growth in highly correlated low participation rates for southern Florida. Western Mississippi shows a contraction of the same behavior. And, eastern Mississippi, along the border with Alabama, shows growth in highly correlated high participation that expands north over time.

Moving to more targeted populations, Figure 5.3B presents Subgroup 1 where we observe similar clustering patterns as the full population. However, the areas of high participation start and remain smaller than the population counterpart. Areas of low participation are more distributed throughout the south but urban clusters are substantially smaller, with the only notable exception of Charleston, West Virginia, and

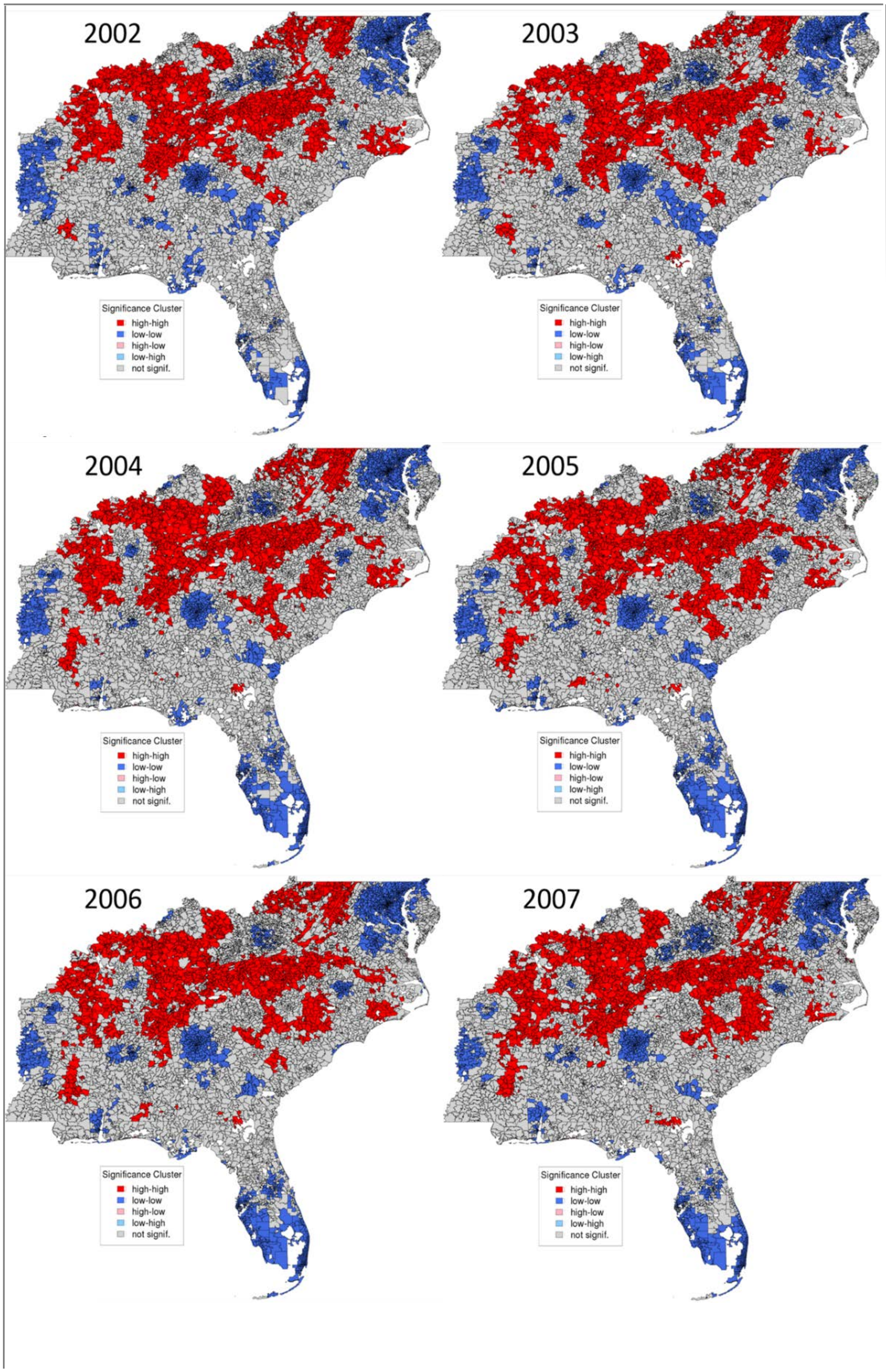
Miami, Florida. These two areas show substantially larger clustering behavior. Areas of high participation are not as concentrated in the west, but larger throughout North Carolina. In fact, for the population, the Raleigh-Durham area is highly correlated low near low participation, but for Subgroup 1 this area flips to highly correlated high near high.

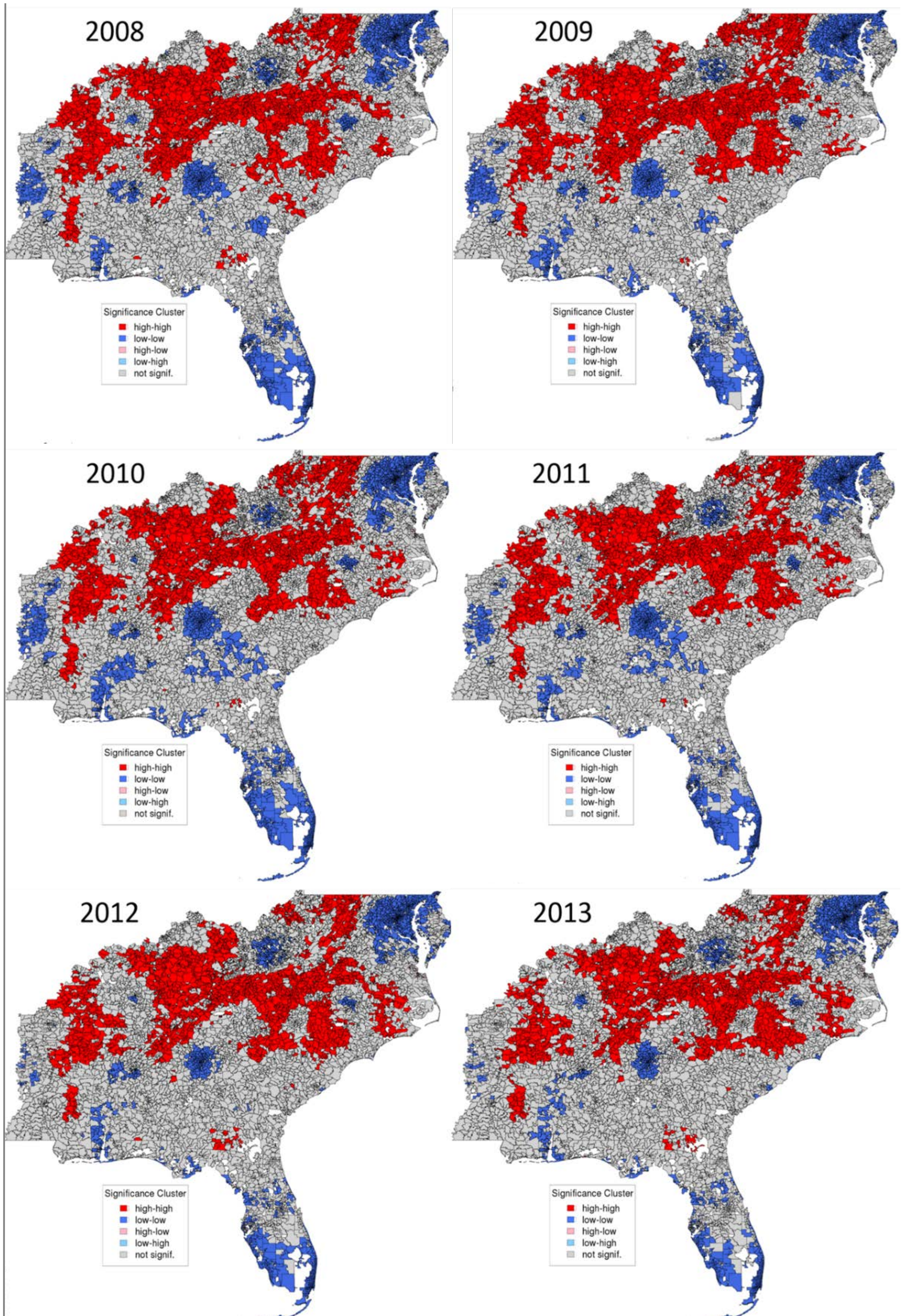
The cluster map for Subgroup 2, in Figure 5.3C, shows the starkest difference from the full population. Initially, Subgroup 2 reports generally the same patterns as before, highly clustered low participation in the south, generally highly clustered high participation in the Appalachian and areas west. And, like Subgroup 1, urban areas are much less concentrated. There is a larger high participation concentration on the North Carolina coast, and newly identified areas appearing in central Florida and Maryland's Eastern Shore. As time passes, the observance of clustering drastically declines, and by 2013 only scattered pockets of spatial autocorrelation remain. They include, most notably, for high participation: the northern border between Virginia and West Virginia, eastern Kentucky and Tennessee, and southern Georgia, and for low participation: western Mississippi, southern Alabama into Atlanta, and D.C.

While these maps provide a good sense of significance clusters and a direct view into the locationality, they do not provide much meaningful information regarding the split between the number of spatially correlated ZIP Codes and not, and the split between the number of areas showing high and low participation clusters. For that, I reorganize this information into distributional clusters between high and low, Figure 5.4A-C. Specifically, I report the percent of ZIP Code concentrations by type of cluster for each year. With these we can more clearly observe which participation cluster dominates, whether these clusters are becoming more concentrated or less, and whether spatial correlation in general is expanding or contracting over time. For all eligibility groups, as would be expected, the distribution of spatial clusters diminishes over time. This reiterates the earlier findings that location becomes less of a factor in Saver's Credit participation across time. As such, this trend offers a glimpse into the dispersion rate of awareness. In general, it was not until 2008 or 2009 that dispersion took off and spatial autocorrelation began to wane.

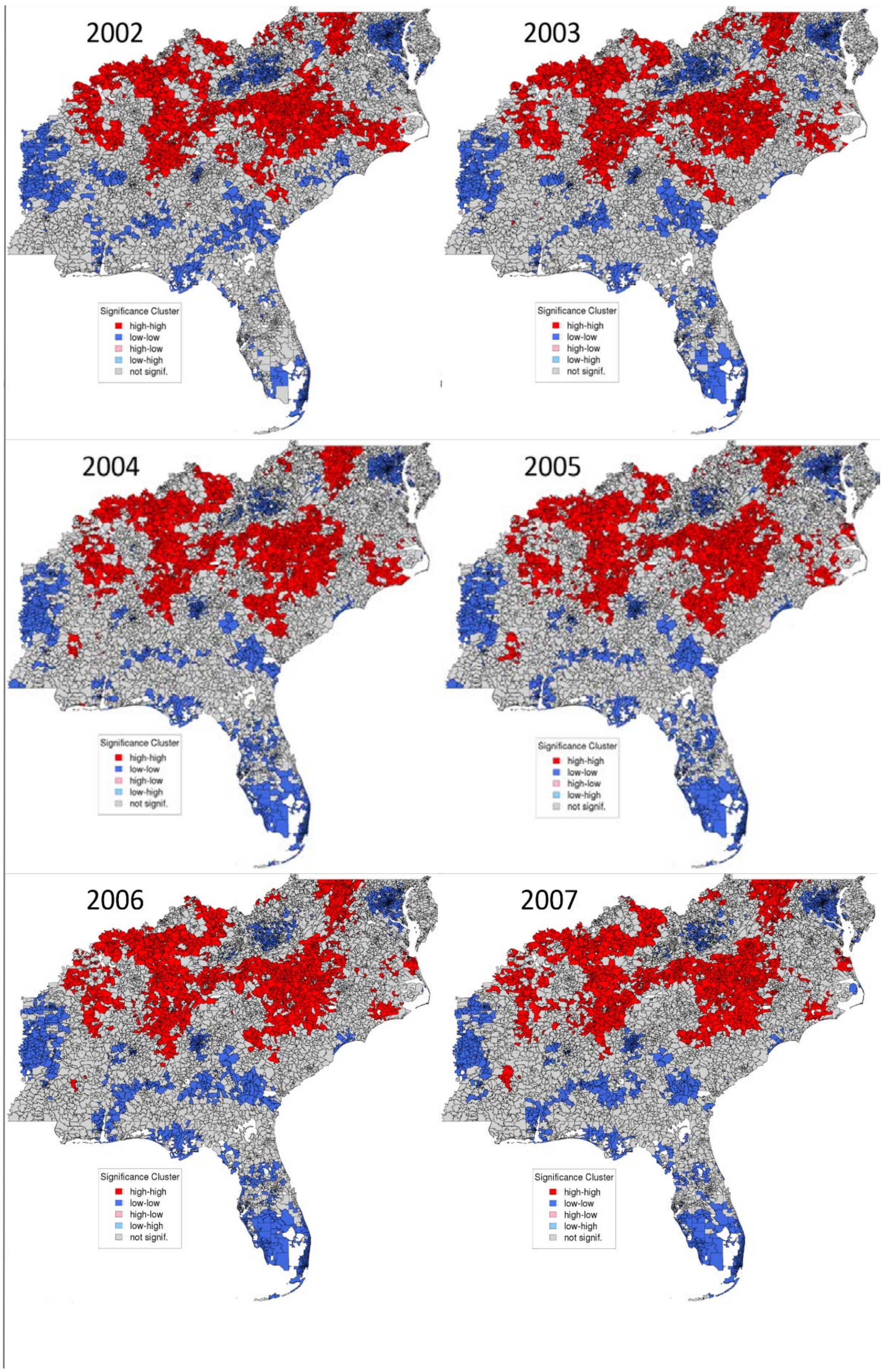
Two unexpected trends emerge from these figures. First, spatial autocorrelation expands in the first few years of Saver's Credit availability. This contradicts expectation, one would expect that location would matter less as knowledge of the credit dispersed. The pattern of becoming more correlated is most apparent for the loosest eligibility definition, the full population, Figure 5.4A. This ostensibly suggests that participation in the Saver's Credit is compressing over time as opposed to dissipating over time. Such findings, from a policy perspective are quite troubling, as it suggests that the knowledge of this credit is not spreading across the population as time goes on, further limiting its usefulness to induce retirement savings among those less inclined to adequately save for retirement. As we target our population of interest to more stringent eligibility definitions, this observed trend becomes less apparent. This could only be the case if populations naturally segregate themselves socioeconomically and in turn strengthen the spatial dependence by having less information sharing between segregated areas. Such behavior has been well documented in the literature (Reardon et al. 2018; Fogli and Guerrieri, 2019; Owens 2019). By Subgroup 2, after controlling for income, tax liability, and retirement contribution activity, the pattern generally returns to expectation.

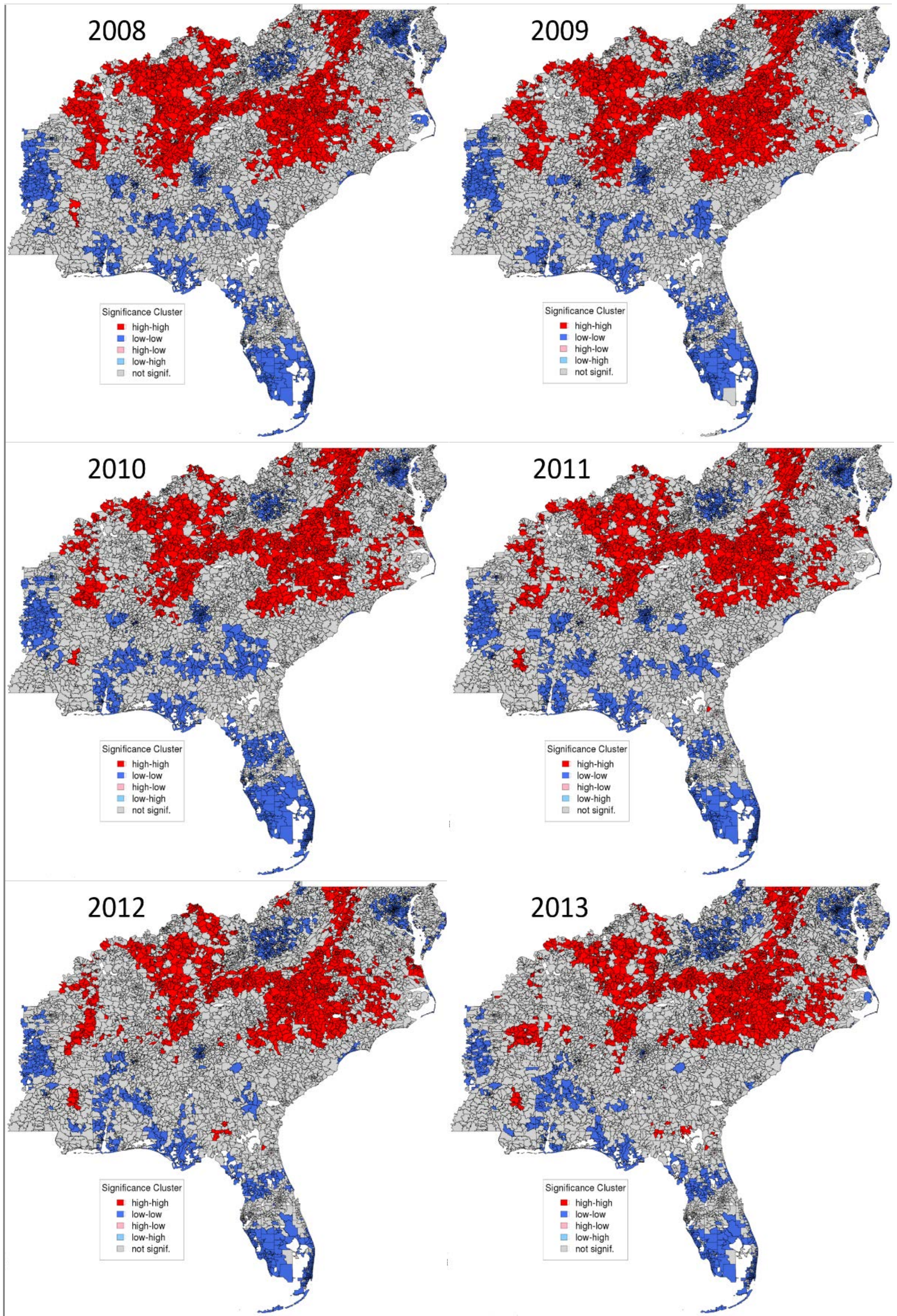
The second is the dominance of low participation clustering over high. For every eligibility group, the proportion of ZIP Codes with low participation correlated with other low participation ZIP Codes is consistently larger than high participation. In fact, the only year in which this is not the case is in Subgroup 2 for 2006, when they are all but equal. The justification is not entirely clear. Knowing that urban areas tend to be more likely to report highly correlated low participation, it would seem to reason that urban areas are either more likely to not be aware of the credit, actively avoid the utilization of the credit, or some combination thereof. If the former, then it would also have to be true that these areas remained relatively isolated from the surrounding rural areas reporting highly correlated high participation. While unlikely, this is not entirely without some merit. It is often the case that urban and suburban areas spend little time in communion with rural areas. However, the underlying mechanism causing this high/low division remains unanswered.


2002
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2003
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2004
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2005
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2006
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2007
Significance Cluster
high-high
low-low
high-low
low-high
not signif.

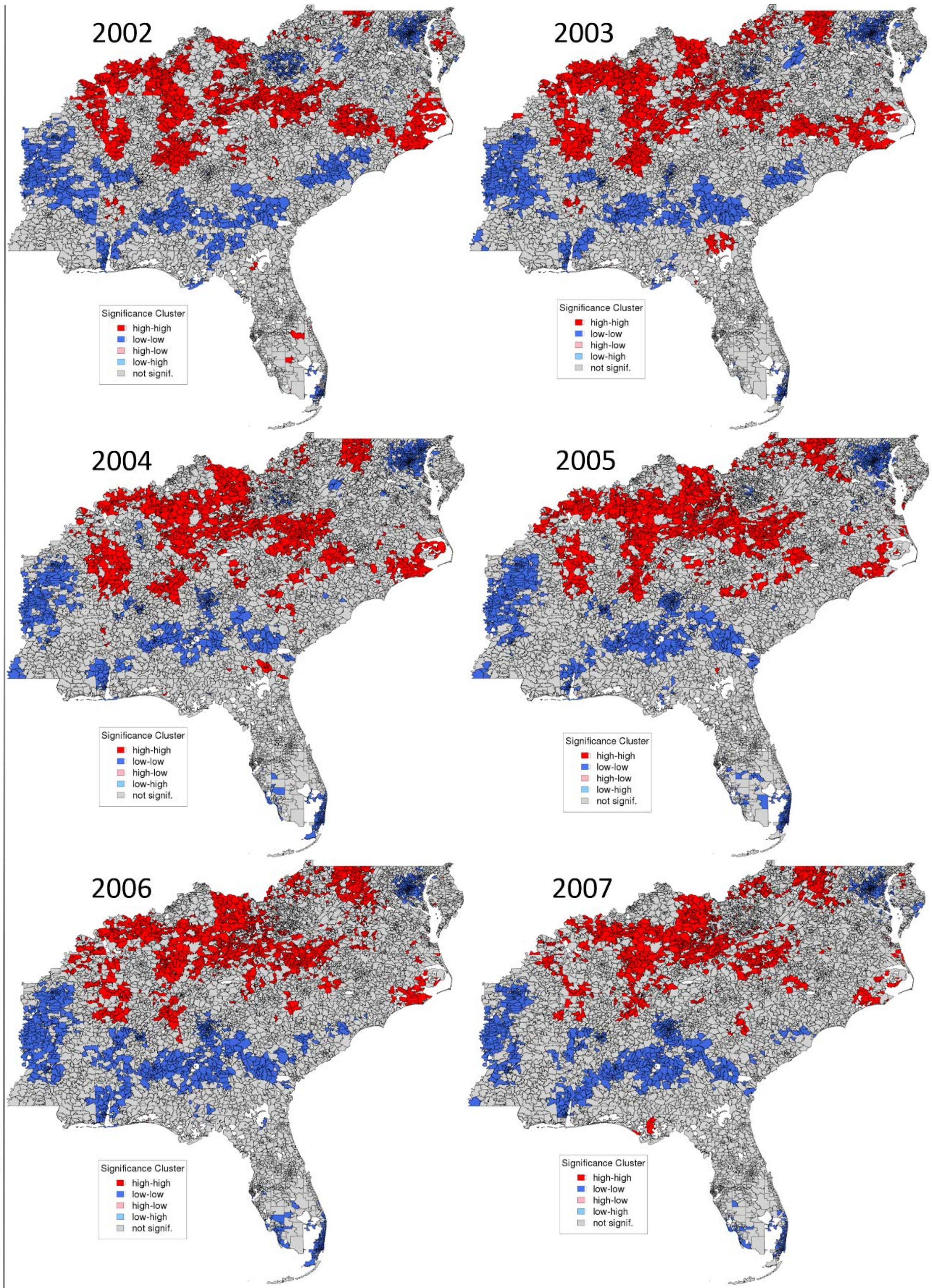


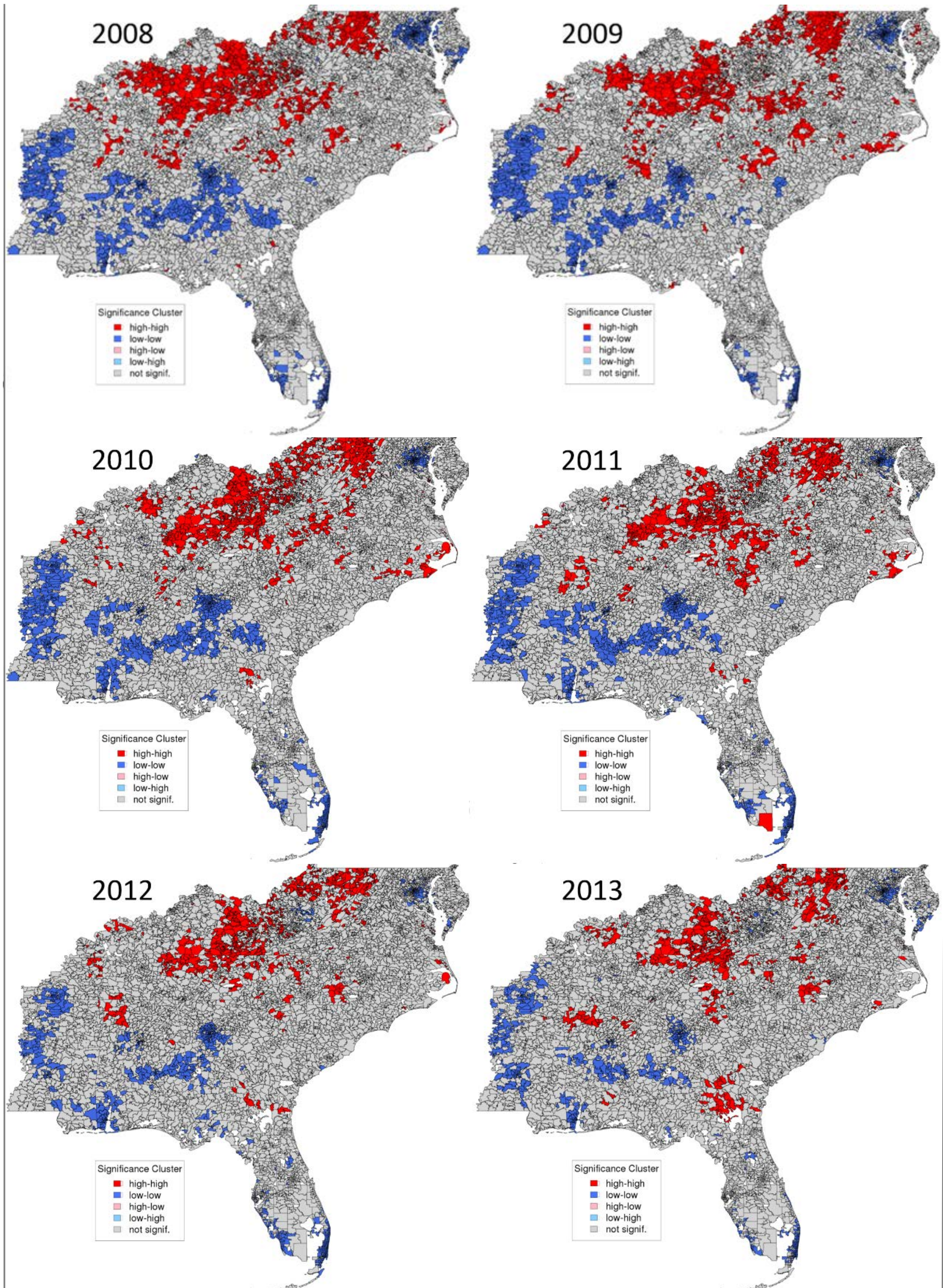
A) Full Population


2002
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2003
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2004
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2005
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2006
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2007
Significance Cluster
high-high
low-low
high-low
low-high
not signif.


2008
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2009
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2010
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2011
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2012
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2013
Significance Cluster
high-high
low-low
high-low
low-high
not signif.

B) Subgroup 1


2002
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2003
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2004
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2005
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2006
Significance Cluster
high-high
low-low
high-low
low-high
not signif.
2007
Significance Cluster
high-high
low-low
high-low
low-high
not signif.



C) Subgroup 3

Figure 5.3: Cluster Map of Saver's Credit Participation by Eligibility Group and ZIP Code, 2002-2013

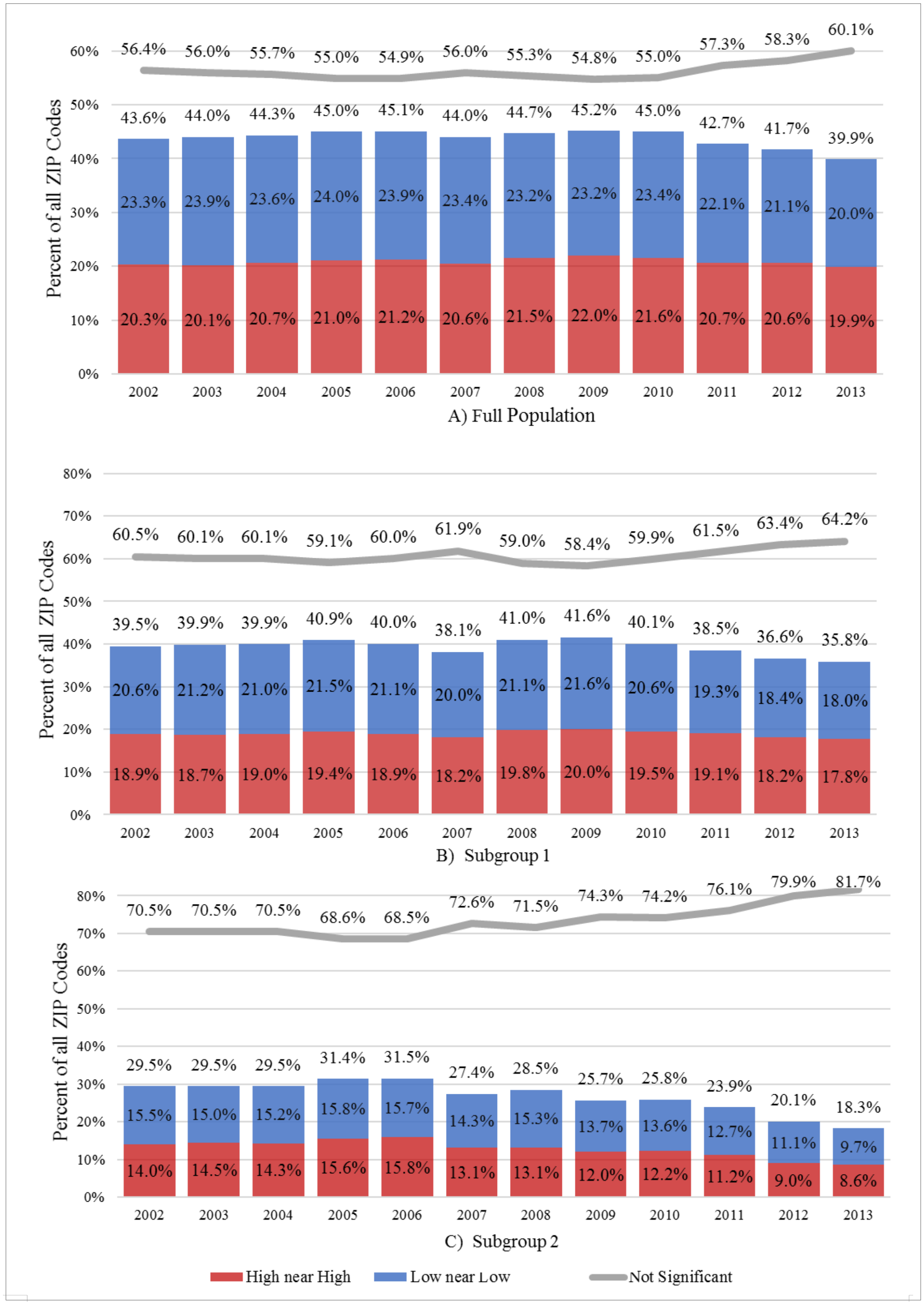


Figure 5.4: Percent of ZIP Code Concentrations by Type of Cluster and Eligibility Group in Southeast Region, 2002-2013

Section 5.4 - Discussion

Through statistically measuring the spatial autocorrelation of the Saver's Credit participation across the U.S. and Southeast region, I have explicitly confirmed the stipulated correlation underlying localities from Chapter 4. Through employment of the Moran's I statistic spatial autocorrelation exists but erodes overtime. These findings were robust to other spatial measurements. Such findings mark the first statistical proof that take-up of the Saver's Credit is at least initially reliant on the location in which someone resides. As we moved toward more targeted sub-populations, and across time, these correlations diminished. When converted to marginal rates of change, I show a momentum that was quite volatile, with relatively large ebbs and flows to the rate of correlation diffusion.

I then identified the particular areas of high correlation by way of the Local Moran's I. There were initially large clusters of highly concentrated behavior, with high participation clusters within the Appalachian region and low participation clusters around urban areas, western Mississippi, and southern Georgia. However, these clusters did dissolve over time for more concentrated definitions of eligibility. The seemingly contradictory pattern between the full population and other sub-populations supports the literature on a growing segregation of socio-economic class in general. And again, these results proved to be robust to other measures. The erosion of initially acute spatial clustering indicates that knowledge, and subsequent take-up, of the Saver's Credit is spreading and points to the growing relevance of this credit, in contrast to earlier research that limited analysis to only as late as 2006. The small declines in Subgroups 1 compared to large declines in Subgroup 2 observed in Figures 5.3B and C beg further investigation into potential causal relationships. Such differences may indicate that the Saver's Credit is more an add-on credit for those already choosing to make retirement contributions than actually inducing any new savings behavior, a question I pick up in Chapter 7.

By fully understanding the spatial autocorrelation underpinning participation rates of the Saver's Credit, we are better able to understand the dynamics and direction for which participation has and will spread. The results of this study definitively conclude that behavior observed in one location is dependent on the observed behavior in another. Such conclusions matter for statistically proving that take-up of a tax expenditure is dependent, at least in part, on where one resides. From a policy perspective, knowing that spatial dependence plays a role in tax expenditure activity is critical for better targeting, outreach, and interpretation. For example, knowing that the Saver's Credit is utilized in more rural than urban areas would suggest that outreach should be targeted more heavily at those urban and surrounding areas. Additionally, understanding correlation patterns over time allow us to better predict future dispersion. As observed through Figures 5.2A and B, and corroborated with Figures 5.4A-C, correlation progressively declines over time and across eligibility groups. From the policy perspective here, one can expect to continue to observe this momentum going forward.

Research in the field of tax expenditures have generally ignored the spatial dynamics underpinning expenditure activity. The analysis in this study can be applied to understanding how other expenditure activity may present itself and spread over time. Where previous research into expenditure activity has tended towards a national perspective, this study exposes the flaw in such an approach. By uncovering the spatial distribution underpinning national estimates of tax expenditure activity, one can better estimate the local impact of a given expenditure.

Further, exposing spatial correlation serves to highlight the interaction between federal tax expenditures and localities. While nearly all federal tax expenditures are available across the country without regard to the location of the filing individual, this study highlights the disparity, and embedded spatial dependence in actual utilization. This implies then, that while tax expenditure opportunities are offered without partiality to location, take-up of those same expenditures is.

CHAPTER 6 - MEASURING THE SPATIAL SPILLOVER EFFECT

The previous study presented conclusive evidence that spatial autocorrelation exists between ZIP Codes with respect to Saver's Credit participation. It also spent a significant amount of time identifying where those areas were located. However, it remained silent on the impact of other filing characteristics that may also be influencing observed participation patterns. It could very well be the case that areas showing high participation are being influenced by having more returns filed by a paid preparer, more heavily concentrated with filers of a particular filing status, or income sources that drives participation. In this next study I employ various spatial regression models to account for the dependence underpinning observed behaviors. This allows for better understanding of what factors contribute to explaining locational participation, along with highlighting and quantifying the spatial spillovers that underpin the observed dependence from Chapter 5.

Specifically, I examine the spatial spillover, or indirect, effects that filing characteristics have on Saver's Credit participation. By way of applying spatial regression models, I explicitly quantify the strength and direction in spatial diffusion that neighboring ZIP Codes exert on each other. Such influence is considered indirect because the activity of one ZIP Code has influence, or spills over, on the activity in another. Another way of thinking about it is that of an externality created by one location affects another. I use three different spatial regression models, the Spatial Lag (SAR), Spatial Error (SEM), and Spatial Durbin (SDM), to tease out these indirect effects for which more conventional regression models ignore.

In application, this study addresses the often-violated assumption in more conventional models, that of independence between observations. Violations here lead to, at the very least, unreliable standard errors, but more often biased estimates, calling into question the reliability of the resulting coefficients. With highly localized data, such is the case here, it is unreasonable to expect complete independence between observations, and as my previous research concluded, it is also born out in the data. Spatial regression modeling resolves this issue by allowing for, and explicitly measuring, dependence between observations.

This line of research proceeds as follows. In the next section, I briefly review the literature on the impact of spillover effects of tax expenditures. Then I discuss the empirical methodology employed in this study and compare the models. Next, I present estimation results. And finally, move into a discussion of the results.

Section 6.1 - Literature Review

There has been an explosion of work in public finance exploring these spatial spillovers in recent years. However, previous research has focused almost exclusively on the spatial interactions between one local jurisdiction to another. They explore spatial interactions through one of three theoretical mechanisms: yardstick competition, tax rate competition, and budget spillovers. The first thread of research examines the idea that voters in one jurisdiction observe choices in another's which in turn informs their decisions on the quality of their own jurisdiction, see Besley and Case (1995), Bordignon et al. (2003) and Revelli (2005).

The second thread relates to revenue generation, specifically through jurisdictional tax rate choices effected by others jurisdictional rate setting, often referred to as tax mimicking, see Heyndels and Vuchelen (1998), Brett and Pinkse (2000), Revelli (2001), and Baskaran (2014). Also, see Brueckner (2003), Rork (2003), and Borck et al. (2007) for more on spillovers through competition of mobile tax bases. Encompassed in this is a huge field of research exploring the indirect effects of housing policy and property tax (Brueckner and Saavedra 2001; Sinai and Gyourko, 2004; DeSilva et al. 2012, and Alm et al. 2016).

The third thread focuses on the spatial effects of jurisdictional spending. For an overview of the literature see Solè-Ollè (2006). Research here most often explores jurisdictional spillovers on welfare and benefits, see Brueckner (1998), Saavedra (2000), Baicker (2005), and Ojede et al. (2018). But some work has been done exploring the crowding out/in effect on the private sector from spatial spillovers of public spending, see Conley and Dix (1999) and Funashima and Ohtsuka (2019).

While research into the spillover effects between jurisdictions has exploded, to date, there has been very little work exploring the local spatial spillover effects of federal tax expenditures. Alm and Yunus (2009) examined the spatial dependence of tax evasion between taxpayers across state boundaries. By also controlling for serial autocorrelation, they found a positive and significant spatial effect on tax evasion and persistence across taxpayers with respect to tax evasion. Albouy (2009) explored the interaction of federal income tax rates on the geographic distribution of workers. Specifically, he empirically measured the difference in federal taxes paid between identical workers located in low-wage verses high-wage cities. He found that workers in high-wage cities pay up to 27 percent more in federal taxes than the same workers in low-wage cities. Chetty et al. (2015) explored how federal tax expenditures impact local intergenerational mobility. They use variations in observed total itemized deductions, mortgage interest deductions, and federal Earned Income Tax Credits at the ZIP Code level to determine their local influence on intergenerational mobility. They found that tax expenditures are positively correlated with higher levels of intergenerational mobility, suggesting that targeted tax expenditures on lower income households would positively, and significantly, effect economic opportunity. As compared to their study, I explore factors that impact federal tax expenditure participation rather than what federal tax expenditure activity impacts.

My work explicitly examines the impact that location has on federal tax expenditure activity, specifically on participation in the Saver's Credit. No other research into this credit has attempted to measure these effects. In fact, all previous work, with the exception of Ramnath (2013), have ignored exploring the factors leading to credit participation. Ramnath, through a probit model, measured the probability of claiming a Saver's Credit conditional on eligibility. Her definition of eligibility was exclusively those that reported income below the cutoff and had a tax liability after other non-refundable credits, equivalently Subgroup 1 of my research. She found participation was positively associated with wage and self-employment income, e-filing, and using a paid preparer, and negatively associated with filing single or married filing separately, relative to filing jointly. She did not include any assessment on locational effect, nor would her dataset have been appropriate to do so[16]. My research expands on this by explicitly measuring spatial effects and expanding the definition of eligibility to include those that also make retirement contributions.

Section 6.2 - Empirical Methodology and Model Specification

The starting point for this analysis is by way of a simple Ordinary Least Squares (OLS) model.

$$Y_t = X_t\beta_t + \varepsilon_t \tag{6.1}$$

where *Y* is a matrix of the percent of eligible population taking the Saver's Credit in each ZIP Code, *i*, in year *t*. *X* represents an $n \times K$ matrix of observable characteristics, where *n* is the total number of ZIP Codes and *K* is the number of explanatory variables. β is a parameter vector and ε is the error term vector. Henceforth, I drop the year subscript but it is assumed for all variables and parameters. Among other assumptions, the OLS model assumes two conditions that will be pertinent to this analysis.

Assumption 6.1: The error has an expected mean of zero.

This states that no observations of *X* convey information about ε. In other words, the error term is independent of *X*.

$$E(\varepsilon|X) = 0 \tag{6.2}$$

Assumption 6.2: Errors are spherical.

[16] Heim and Lurie (2014) control for region in their Difference in Difference model but do not include those estimates in their reported results, nor do they measure the determinants of Saver's Credit participation but rather the probability of making a retirement contribution conditional on being eligible for the credit based on income level.

This speaks to the variance-covariance matrix and embeds two conditions. First, the variance between X and ε is constant, or homoscedastic. Second, the covariance of ε is zero, or nonautocorrelated.

$$E(\varepsilon\varepsilon'|X) = \sigma^2 I_n \tag{6.3}$$

With these assumptions β is correctly estimated as

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{6.4}$$

$$Var(\widehat{\beta}) = \sigma^2(X'X)^{-1} \tag{6.5}$$

However, when these assumptions are violated $\hat{\beta}$ is incorrectly estimated to be either biased, as is the case when Assumption 6.1 is violated, or have inefficiently estimated standard errors, as is the case when Assumption 6.2 is violated.

Section 6.2.1 - Spatial Lag Model

If we assume there exists spatial spillovers of credit participation across ZIP Codes, as is implied by Chapters 4 and 5 above, then the conventional OLS model, taking the form of equation (6.1), throws this recursive relationship into the error term. Without directly accounting for this in the model, the assumed specification attributes a portion of this unmeasured relationship to the measured X and biases the estimate of $\hat{\beta}$. This can be shown by first starting with the OLS model, but replacing ε with u.

$$Y = X\beta + u \tag{6.6}$$

$$u = \rho WY + \varepsilon \tag{6.7}$$

where WY is the endogenous lag variable for the spatial weight matrix W. We can now rewrite equation (6.7)

$$Y = X\beta + \rho WY + \varepsilon \tag{6.8}$$

And solving for β

$$(I_n - \rho W)Y = X\beta + \varepsilon \qquad (6.9)$$

$$Y = (I_n - \rho W)^{-1}X\beta + (I_n - \rho W)^{-1}\varepsilon \qquad (6.10)$$

$$\hat{\beta} = (X'X)^{-1}X'(I_n - \rho W)Y \qquad (6.11)$$

$$\hat{\beta}(I_n - \rho W)^{-1} = (X'X)^{-1}X'Y \qquad (6.12)$$

We see that OLS is biased by size $(I_n - \rho W)^{-1}$. Here, ρ measures the spatial dependence, or spillover, between observation *i* and *j*. Notice that (6.8) resolves to (6.1) when ρ = 0, so if spatial spillovers are not present, then the OLS model would still hold in theory. Equation (6.8) defines the spatial autogressive model, as I will refer to as the Spatial Lag Model, (SAR). By construction, this model is akin to an AR(1) model for time series analysis.

Before moving on, there is still the matter of articulating the interpretation of the derived coefficient(s). When ρ = 0 the left-hand side of equation (6.12) for the K^{th}exogenous variable resolves to

$$\hat{\beta}_K(I_n - \rho W)^{-1} = \hat{\beta}_K I_n \qquad (6.13)$$

Following Golgher and Voss (2016) let us assume a simple 3 × 3 spatial weight matrix taking the form

$$W = \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix} \qquad (6.14)$$

when $\rho \neq 0$ the left-hand side of equation (6.12) resolves to

$$\hat{\beta}_K (I_n - \rho W)^{-1} = \left(\frac{\hat{\beta}_K}{1-\rho^2}\right) \begin{pmatrix} (1-\rho^2) & \rho & \rho^2/2 \\ \rho/2 & 1 & \rho/2 \\ \rho^2/2 & \rho & \left(1 - \rho^2/2\right) \end{pmatrix} \quad (6.15)$$

This demonstrates that the estimated β from equation (6.8) is only part of the larger picture of effect, and is intrinsically dependent on ρ. Therefore, the direct effect (DE), equal to β in the OLS model, is no longer merely β in the SAR but rather the mean of the diagonal terms, represented in this simple 3 × 3 example as:

$$DE = \left(\frac{1}{3}\right)\left(\frac{\hat{\beta}_K}{1-\rho^2}\right)\left[(1-\rho^2) + 1 + \left(1 - \rho^2/2\right)\right] \quad (6.16)$$

And the off-diagonal elements represent the indirect effects (IE), or spillovers, with the average IE being the represented here by

$$IE = \left(\frac{1}{3}\right)\left(\frac{\hat{\beta}_K}{1-\rho^2}\right)\left[\rho/2 + \rho^2/2 + \rho + \rho + \rho^2/2 + \rho/2\right] \quad (6.17)$$

Section 6.2.2 - Spatial Error Model

If spatial dependence was being driven by other factors not captured through a recursive relationship, spatial autocorrelation would be present in the error term. This results in a violation of Assumption 6.2 with the error term not being independent and the expected variance-covariance matrix not resolving to $\sigma^2 I_n$ but rather to some other positive definite matrix, Ω

$$E(\varepsilon\varepsilon'|X) = \sigma^2\Omega \quad (6.18)$$

When such an autocorrelation exists, the estimate of β remains unbiased, but is no longer efficient. To see this, we turn back to equation (6.6) but now u is redefined as

$$u = \lambda W u + \varepsilon \quad (6.19)$$

$$u = (I - \lambda W)^{-1} \varepsilon \quad (6.20)$$

And results in a variance-covariance matrix for u

$$E(uu'|X) = \sigma^2 [(I - \lambda W)'(I - \lambda W)]^{-1}$$

$$= \sigma^2 \Omega \quad (6.21)$$

Then, similar to the generalized least squares estimator[17], $\hat{\beta}$ resolves to

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \quad (6.22)$$

Thus, the estimate of β may remain unbiased but the variance of β is no longer efficient. This is more explicitly seen when looking at the variance equation.

$$Var(\widehat{\beta}) = \sigma^2 (X'\Omega^{-1}X)^{-1} \quad (6.23)$$

OLS thus underestimates the variance of $\hat{\beta}$ by assuming Ω^{-1} is equal to 0. This results in the OLS model losing its attractiveness given this potential for inconsistent standard errors. We can rewrite equation (6.6)

$$Y = X\beta + (I - \lambda W)^{-1} \varepsilon \quad (6.24)$$

$$Y = X\beta + \lambda W u + \varepsilon \quad (6.25)$$

Equations (6.24) and (6.25) are equivalent and define the Spatial Error Model (SEM).

Unlike the SAR the interpretation of the derived coefficient(s) is the same as that for OLS. This is because the spatial weight matrix only impacts the error term, leaving the expected value of the dependent variable untouched.

[17] See Greene (2003) for the complete derivation.

Section 6.2.3 - Spatial Durbin Model

In the previous example, spatial autocorrelation was shown to affect the efficiency of the OLS model. However, it could well be the case that autocorrelation in the error term is masking the more serious issue of omitting a relevant variable, independent of X but unspecified. In such a situation the estimated β is no longer just inefficient but biased as well.

The SEM assumes autocorrelation is due only to an autoregressive process in the error term. It may well be the case that the spatial residual autocorrelation is capturing a correlation between ε and X. This by definition violates Assumption 6.1 and leads to a biased $\hat{\beta}$. We can model the strength of this relationship with

$$z = X\gamma + \varepsilon \tag{6.26}$$

where the scalar γ and variance, σ^2, of the error term determine the strength. Since we have already established the error follows a spatial autoregressive process, we simply redefine u from equation (6.19) to be

$$u = (I - \lambda W)^{-1} z \tag{6.27}$$

We can now rewrite equation (6.6) to account for both the autocorrelation in the error term and the relationship between X and ε.

$$Y = X\beta + (I - \lambda W)^{-1}(X\gamma + \varepsilon) \tag{6.28}$$

$$Y = X\beta + (I - \lambda W)^{-1}\gamma X + (I - \lambda W)^{-1}\varepsilon$$

$$(I - \lambda W)Y = (I - \lambda W)X\beta + \gamma X + \varepsilon$$

$$Y = \lambda WY + X(\beta + \gamma) + WX(-\lambda\beta) + \varepsilon \tag{6.29}$$

Notice if we convert λ to ρ then the first term in equation (6.25) is the same as the second term in equation (6.8). If we also set θ = $(-\lambda\beta)$ and add together the scalar terms β and γ

$$Y = \rho WY + X\beta_{new} + \theta WX + \varepsilon \tag{6.30}$$

Ignoring this type of interaction violates both Assumptions 6.1 and 6.2. Assumption 6.1 by way of both omitting γ from the calculation of β and ignoring ρWY all together. Assumption 6.2 by way of throwing X and ε interactions into the error term. Equations (6.29) and (6.30) define the Spatial Durbin Model (SDM). Here, there are two additional parameters from the traditional OLS model of Equation (6.1), one for lag of X and one for lag Y in the form of ρ and θ respectively. Table 6.1 lays out the four models[18].

Table 6.1: Description of Models

Model	Specification	Spatial Parameter	Spatial Description
Ordinary Least Squares, OLS	$Y = X\beta + \varepsilon$		No spatial interaction
Spatial Lag, SAR	$Y = \rho WY + X\beta + \varepsilon$	ρ	Spatially endogenous interactions
Spatial Error, SER	$Y = X\beta + \lambda Wu + \varepsilon$	λ	Spatial interactions in the error
Spatial Durbin, SDM	$Y = \rho WY + X\beta + \theta WX + \varepsilon$	ρ,θ	Spatially endogenous and spatially lagged exogenous interactions

Much like the SAR, the interpretation of the derived coefficients from the SDM is not immediately intuitive. Solving for β for the K^{th}exogenous variable in equation (6.30) resolves to

$$[I - \rho W]^{-1}\left[\widehat{\beta_K}\, I + \theta_K W\right] = (X'X)^{-1}X'Y \tag{6.31}$$

[18] For a fuller discussion, see LeSage and Pace (2009) and Le Gallo (2014).

The difference between equation (6.31) for the SDM and (6.12) for the SAR is the addition of the spatially dependent independent variable captured in $I + \theta_K W$. Using the illustrative spatial weight matrix example from (6.14), the left-hand side of (6.31) resolves to

$$[I - \rho W]^{-1}[\widehat{\beta_K}\, I + \theta_K W] = \left(\frac{1}{1-\rho^2}\right)\begin{pmatrix} (1-\rho^2) & \rho & \rho^2/2 \\ \rho/2 & 1 & \rho/2 \\ \rho^2/2 & \rho & \left(1-\rho^2/2\right) \end{pmatrix}\begin{pmatrix} \hat{\beta}_K & \theta_K & 0 \\ \theta_K/2 & \hat{\beta}_K & \theta_K/2 \\ 0 & \theta_K & \hat{\beta}_K \end{pmatrix} \quad (6.32)$$

$$[I - \rho W]^{-1}[\widehat{\beta_K}\, I + \theta_K W] = \left(\frac{1}{1-\rho^2}\right)\begin{pmatrix} (1-\rho^2)\hat{\beta}_K + \frac{\rho\theta_K}{2} & \rho\hat{\beta}_K + \theta_K & \frac{\rho^2\widehat{\beta}_K}{2} + \frac{\rho\theta_K}{2} \\ \frac{\rho\widehat{\beta}_K}{2} + \frac{\theta_K}{2} & \hat{\beta}_K + \rho\theta_K & \frac{\rho\widehat{\beta}_K}{2} + \frac{\theta_K}{2} \\ \frac{\rho^2\widehat{\beta}_K}{2} + \frac{\rho\theta_K}{2} & \rho\hat{\beta}_K + \theta_K & \left(1-\rho^2/2\right)\hat{\beta}_K + \frac{\rho\theta_K}{2} \end{pmatrix} (6.33)$$

The direct effect is once again the mean of the diagonal elements:

$$DE = \frac{3-\rho^2}{3(1-\rho^2)}\hat{\beta}_K + \frac{2\rho}{3(1-\rho^2)}\theta_K \quad (6.34)$$

And the average indirect effect is the mean of the off-diagonal elements

$$IE = \frac{3\rho-\rho^2}{3(1-\rho^2)}\hat{\beta}_K + \frac{3+\rho}{3(1-\rho^2)}\theta_K \quad (6.35)$$

Given the geographic granularity of my data, it is reasonable to assume there is some spatial dependence embedded, not to mention the results of Chapters 4 and 5 suggesting as such. When spatial dependence is not accounted for in the model, as we have just seen, the resulting coefficient estimates are unreliable at best. Having thus described the potential error in applying an OLS model, and deriving three alternative models, I now move on to applying each.

Section 6.3 - Empirical Results

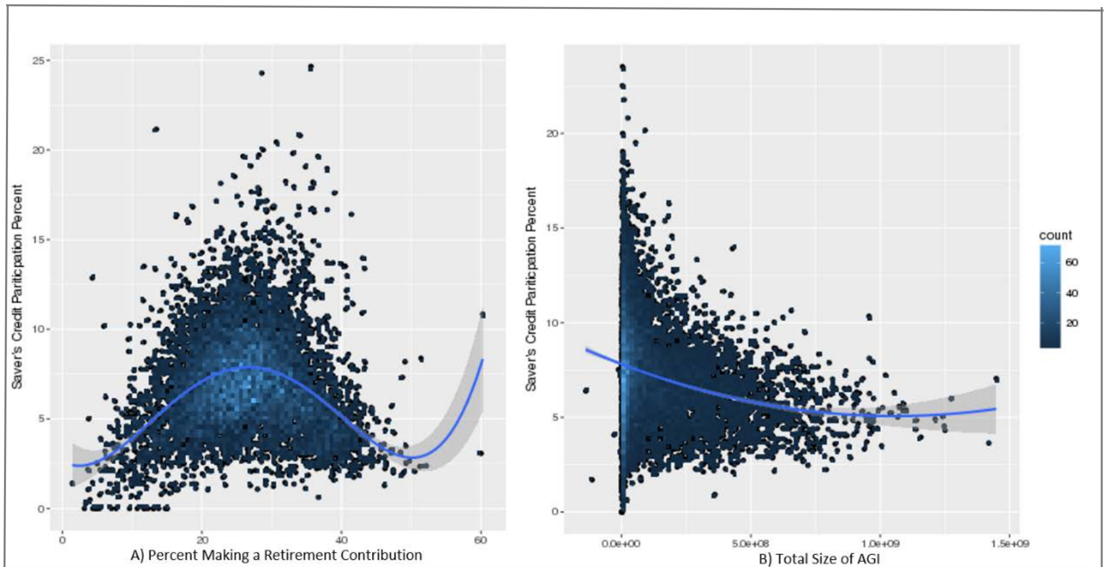
For this research I focused exclusively on information obtainable from IRS administrative data systems[19]. The dependent variable for the models is the percent of taxpayers claiming the Saver's Credit in a given ZIP Code. Table 6.2 lists out each of my explanatory variables. Notice they align closely with those reported in Table 3.2 with a few exceptions; I add a few additional fields to include the size of adjusted gross income and a fuller break out of age and dependents. For the latter two, I additionally include controls for the percent between the ages of 25 and 34, between 55 and 59, and percent reporting between one and three dependents. I also exclude some variables to reduce multicollinearity concerns. These include percent qualifying for 50 percent credit, percent having 50 percent or more of income from self-employment, percent having either 50 percent or 90 percent or more of income from wages, and percent single.

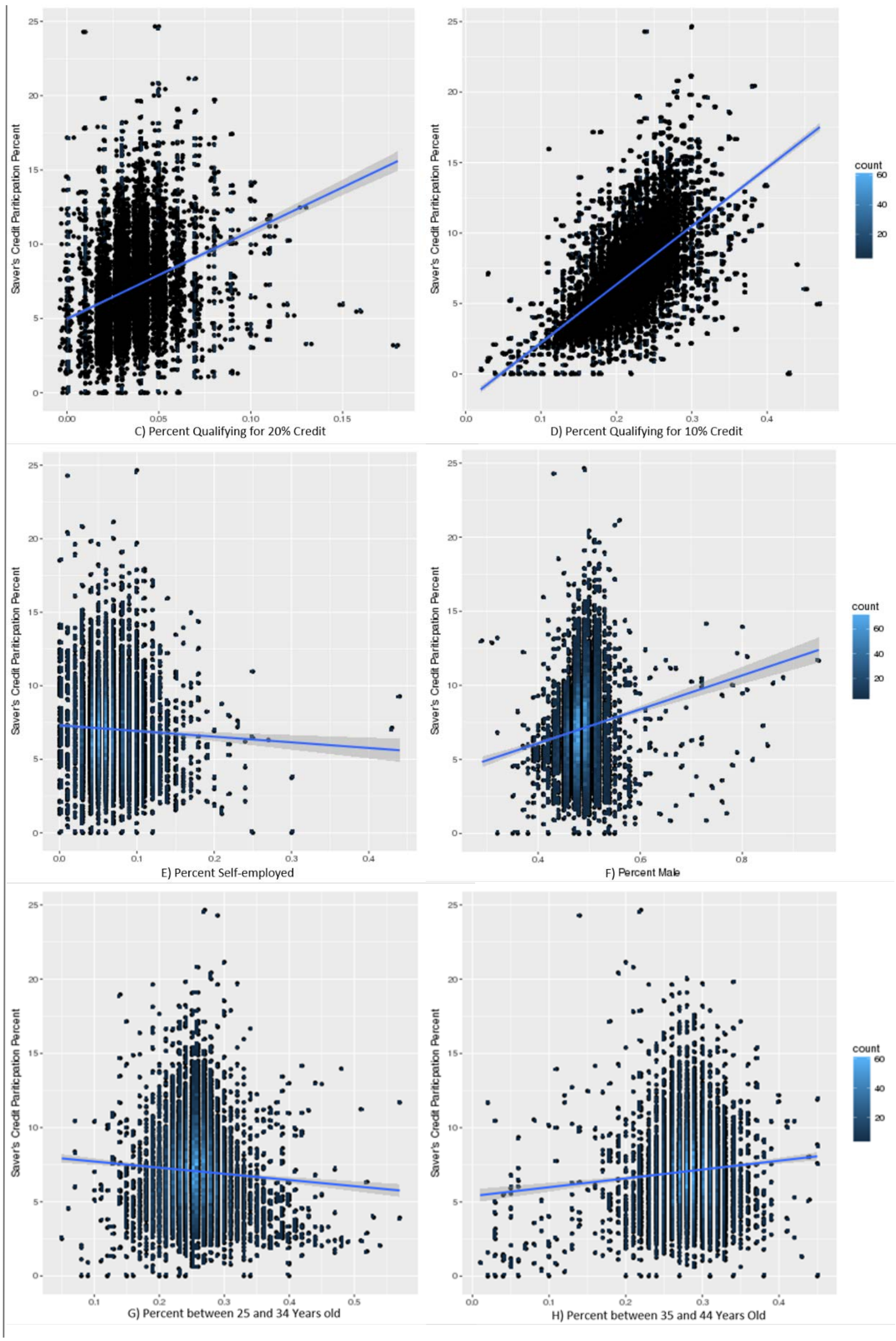
Table 6.2: List of Explanatory Variables

Percent making a retirement contribution
Squared percent making a retirement contribution
Size of AGI
Squared size of AGI
Percent qualifying for 20% credit
Percent qualifying for 10% credit
Percent with 90% or more of income from Self-employment
Percent Male
Percent between 25 and 34 years old
Percent between 35 and 44 years old
Percent between 45 and 54 years old
Percent between 55 and 59 years old
Percent married filing jointly
Percent filing head of household
Percent with 1 to 3 dependents
Percent with 4 or more dependents
Percent using a paid preparer
Squared percent using a paid preparer

[19] See Appendix C.

Because these four models are linear regressions, it is important to first examine the underlying relationship between the percent claiming the Saver's Credit and each parameter variable. Figures 6.1A-N are a sampling of scatterplots pulled from the full population of just that. The percent claiming the Saver's Credit from each ZIP Code is on the y-axis and the various explanatory variables are on the x-axes. I have overlaid an estimated linear model on each to better highlight the observed relationships. Notice however that making a retirement contribution, size of AGI, and percent using a paid preparer all appear nonlinear. Therefore, I include both linear and squared terms for these inputs in my models. Lastly, I standardize all inputs by dividing by the variables' standard deviation. This not only assists in limiting multicollinearity but allows for better comparisons between parameters, to aide in identifying the most influential parameters in the model.




Saver's Credit Paritcipation Percent
C) Percent Qualifying for 20% Credit
D) Percent Qualifying for 10% Credit
E) Percent Self-employed
F) Percent Male
G) Percent between 25 and 34 Years old
H) Percent between 35 and 44 Years Old
count

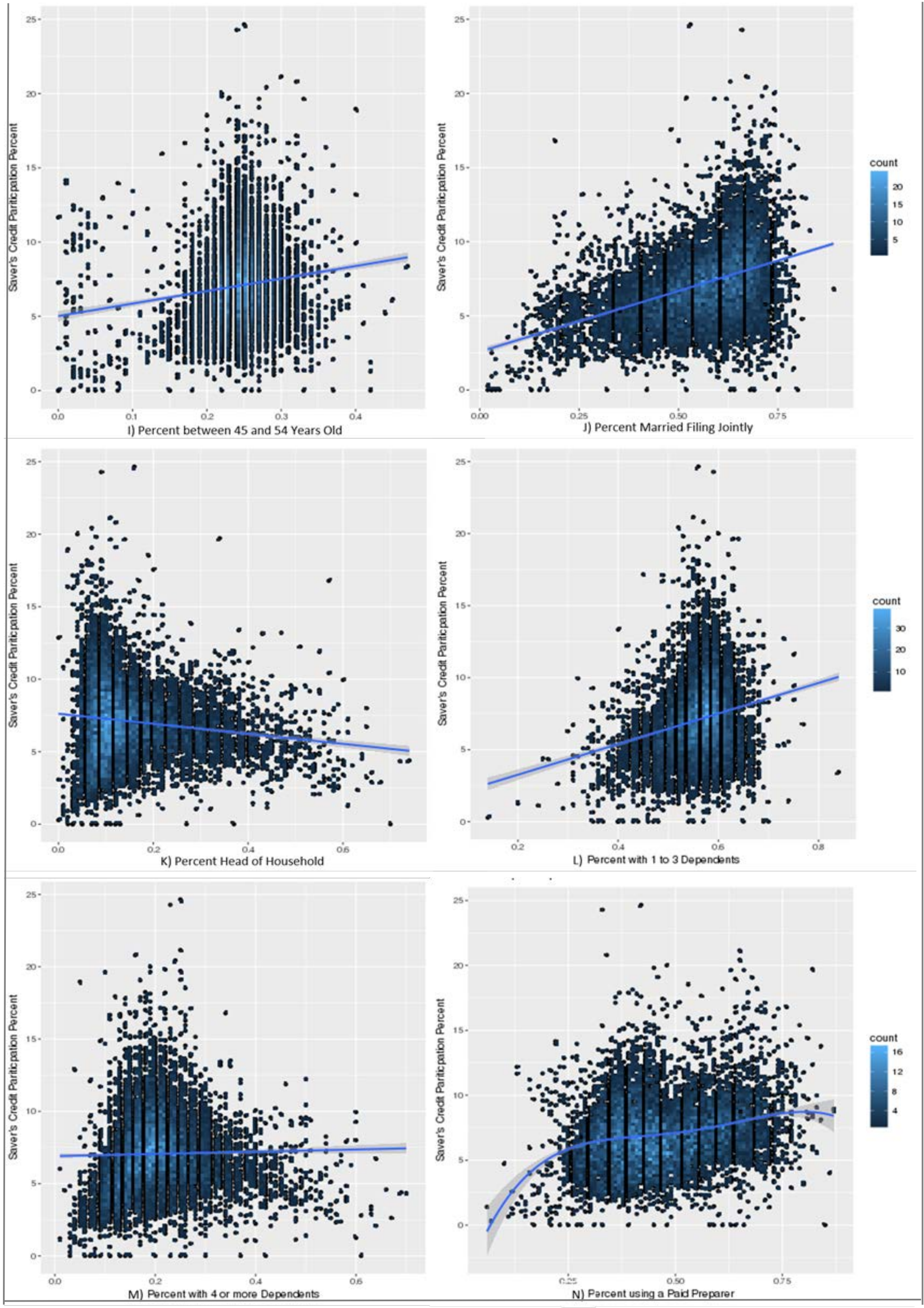

Saver's Credit Paritcipation Percent
I) Percent between 45 and 54 Years Old
J) Percent Married Filing Jointly
count
K) Percent Head of Household
L) Percent with 1 to 3 Dependents
M) Percent with 4 or more Dependents
N) Percent using a Paid Preparer

Figure 6.1: Relationship between Explanatory Variables and Saver's Credit Participation Percent for the Southeast Region, 2002

Section 6.3.1 - Full Population

Full population model estimates are presented in Table 6.3.[20] Column one presents results from the OLS model. It is unsurprising to find that the percent making a retirement contribution has the largest impact on claiming the credit and shows diminishing returns. More surprising is the nearly equally offsetting effect of using a paid preparer, again with diminishing returns, and the negative effect of AGI. Age demographics appear inconsistent with those between 25 and 34 having positive, 35 to 44 negative, and 45 to 59 positive. Self-employment has a minimally positive effect on observed percentages of claiming the Saver's Credit. And qualifying for either the 10 percent or 20 percent credit rate has a positive effect.

When we look at model diagnostics it becomes clear that the OLS is not an appropriate fit. First, applying the Moran's I computation from Chapter 5 to model residuals indicates that spatial correlation is present in the error term. This necessarily indicates that there is spatial dependence observed in the residuals, violating Assumption 6.2 above. Therefore, we can reject this model. Another diagnostic, the Lagrange Multiplier (LM) test, also confirms the presence of spatial dependence. The LM test goes beyond the Moran's I by applying a weighted regression of lagged residuals on the original model residuals (Anselin, 1988). This test can also provide some insight into which spatial dependency model would be more appropriate. All LM tests reject the null

[20] Estimates for each year can be found in Appendix D.

hypothesis of no spatial correlation, strongly concluding that some correction for spatial dependence should be applied. Given the significance of all four tests, it is unclear exactly which spatial model is preferred. While these results may not be conclusive on which model to naturally turn, they necessarily all point to the rejection of the OLS.

Results for the three spatial models are presented in the remaining three columns of Table 6.3. The parameter estimates for the SAR, SEM, and SDM all report the same sign and general magnitude. But in each case, the parameter estimate capturing spatial association, ρ in the case of SAR and SDM and λ in the case of SER, indicates a strong spatial association. When broken down by individual tax year this pattern persists, barring years 2012 and 2013 for the SDM model[21].

Looking at model diagnostics, we can ignore SAR from further investigation. Examining the Moran's I for model residuals parallels that of the OLS. While smaller than before, the Moran's I is still greater than zero, indicating that spatial autocorrelation has not been fully captured through $\hat{\rho}$. The computed Akaike's Information Criteria (AIC), a statistic that measures the comparative quality between econometric models, further supports the preference away from SAR towards either SEM or SDM.

Turning to the latter two models, in both instances the Moran's I confirms the absence of spatial autocorrelation in the residuals and the computed AICs are sufficiently similar. Taking a step back for a moment, the intuition behind these two models may help direct attention to which model is more appropriate. The SEM throws all spatial dependences into the error term and assumes no spillover effects for any observables,

[21] See Appendix D.

Table 6.3: Full Population Model Estimates [1][2]

Parameter [3][4]	OLS	SAR	SER	SDM
Intercept	0.0000	0.0069	0.0821	0.0083
	(0.0070)	(0.0064)	(0.0769)	(0.0062)
Making Retirement Contributions	0.6872	0.5151	0.6415	0.6481
	(0.0409) ***	(0.0379) ***	(0.0390) ***	(0.0390) ***
Squared(Making Retirement Contributions)	-0.3884	-0.2421	-0.2682	-0.2692
	(0.0408) ***	(0.0376) **	(0.0382) ***	(0.0383) ***
Total Adjusted Gross Income	-0.1398	-0.1277	-0.1453	-0.1442
	(0.0192) ***	(0.0176) ***	(0.0181) ***	(0.0181) ***
Squared(Total Adjusted Gross Income)	0.0518	0.0615	0.0695	0.0701
	(0.0171) **	(0.0157) **	(0.0156) **	(0.0156) **
Qualifying for 20 Percent Credit	0.1252	0.0942	0.0952	0.0952
	(0.0079) ***	(0.0073) ***	(0.0071) ***	(0.0071) ***
Qualifying for 10 Percent Credit	0.4451	0.3723	0.3855	0.3826
	(0.0085) ***	(0.0080) ***	(0.0082) ***	(0.0082) ***
90 percent or more of income from SE [5]	0.0422	0.0402	0.0298	0.0293
	(0.0081) **	(0.0075) **	(0.0087) *	(0.0087) *
Male	0.0355	0.0412	0.0463	0.0478
	(0.0092) **	(0.0084) **	(0.0083) **	(0.0083) **
Between 25 and 34	0.0167	0.0244	-0.0021	-0.0032
	(0.0164) *	(0.0150) *	(0.0153) *	(0.0153) *
Between 35 and 44	-0.0134	0.0032	-0.0231	-0.0219
	(0.0108) ***	(0.0099) **	(0.0102) *	(0.0102) *
Between 45 and 54	0.0019	0.0051	-0.0224	-0.0222
	(0.0122) **	(0.0112) *	(0.0114) **	(0.0114) **
Between 55 and 59	-0.0020	-0.0118	-0.0251	-0.0254
	(0.0126) **	(0.0116) *	(0.0115) *	(0.0115) *
Married filing Jointly	0.4301	0.2431	0.2133	0.2048
	(0.0266) ***	(0.0248) ***	(0.0257) ***	(0.0257) ***
Head of Household	0.1977	0.1284	0.1669	0.1658
	(0.0315) ***	(0.0289) **	(0.0306) **	(0.0306) **
1 to 3 dependents	-0.0819	-0.0417	-0.0443	-0.0411
	(0.0154) ***	(0.0141) *	(0.0145) *	(0.0145) *
4 or more dependents	0.0658	0.0709	0.0917	0.0942
	(0.0239) *	(0.0219) *	(0.0232) *	(0.0232) *
Using a paid preparer	-0.4095	-0.1347	-0.2222	-0.2352
	(0.0533) ***	(0.0496) .	(0.0600) *	(0.0605) *
Squared(Using a paid preparer)	0.4504	0.1688	0.3396	0.3508
	(0.0528) ***	(0.0493) *	(0.0569) ***	(0.0574) ***
ρ		0.4992		0.4556
		(0.0123) ***		(0.1288) **
λ			0.8404	
			(0.0119) ***	

Footnotes at the end of table.

Table 6.3: Full Population Model Estimates, cont.

	OLS	SAR	SER	SDM
Lag(Making Retirement Contributions)				-0.4606
				(0.1638) **
Lag(Squared(Making Retirement Contributions))				0.3274
				(0.1655) .
Lag(Total Adjusted Gross Income)				0.1517
				(0.0788) .
Lag(Squared(Total Adjusted Gross Income))				-0.1033
				(0.0823)
Lag(Qualifying for 20 Percent Credit)				0.0282
				(0.0514)
Lag(Qualifying for 10 Percent Credit)				-0.2392
				(0.0407) ***
Lag(90 Percent or more of income from SE[4])				0.0462
				(0.0228) .
Lag(Male)				0.0039
				(0.0474)
Lag(Between 25 and 34)				-0.0482
				(0.0730) .
Lag(Between 35 and 44)				0.0411
				(0.0438)
Lag(Between 45 and 54)				0.0070
				(0.0553) .
Lag(Between 55 and 59)				0.0445
				(0.0625)
Lag(Married filing Jointly)				-0.0843
				(0.0879)
Lag(Head of Household)				-0.1456
				(0.1022) .
Lag(1 to 3 dependents)				0.0535
				(0.0535) .
Lag(4 or more dependents)				-0.0129
				(0.0784)
Lag(Using a paid preparer)				0.4095
				(0.1367) *
Lag(Squared(Using a paid preparer))				-0.5183
				(0.1360) **
N	89,628	89,628	89,628	89,628
AIC	14,878	13,578	13,205	13,122
Moran I Residuals	0.1388 ***	0.0438 ***	(0.0037)	(0.0004)
Log Likelihood		-6,767.81	-6,581.21	-6,522.20
Lagrange Multiplier				
LMError	5,563.47 ***			
LMlag	2,194.87 ***			
RLMerr	3,906.68 ***			
RLMlag	538.08 ***			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Standard Errors in Paretheses below parameter estimates

[1] Population reflects tax filers between 18 and 57 in 1999 with income below $100,000.

[2] Averaged across 2002-2013

[3] All parameters reflect the percent for a given ZIP Code except Total AGI, which is the sum of reported AGI

[4] Lagged parameters are represented by θ in the SDM spatial model.

[5] SE - Self-Employed

Source: author's computation

contained entirely in λ. It is wholly unreasonable to assume no spatial spillovers exist between the explanatory variables. To ignore this fact necessarily introduces an omitted variable bias that can be easily remedied by rejecting the SER in favor of the SDM.

What has only been discussed in abstract is exactly how to interpret the estimated spatial parameters produced by the SDM. The model's interpretation is complicated by its autoregressive structure in both the dependent and independent variables. By allowing ZIP Codes *i* and *j* to depend on each other simultaneously, any change in *i* leads to a change in *j*, which simultaneously results in a change in *i*. We allow this spillover to occur not just in our dependent, but in all observable variables. Therefore, the observed coefficients in Table 6.3 are only part of the full derivation of impact. As presented in equations (6.32), (6.33), and (6.34) the produced coefficient from the regression model cannot be taken on its face as the direct effect. Notice in the derivation of the direct effect, equation (6.33) for a 3×3 weight matrix, the computed $\hat{\beta}$ is adjusted by both the size of the matrix and the size of $\hat{\rho}$.[22]

Table 6.4 presents the direct, indirect, and total effects for the SDM model. The direct effect is the expected impact on the percent participating in the Saver's Credit given a one standard deviational change in the given parameter. The indirect effect measures the spillover across ZIP Codes. This represents the expected impact on participation given a change in a neighboring ZIP Code's explanatory variable. From equation (6.34) representing a 3×3 weight matrix, the indirect effect depends on the

[22] The exact derivation is extensively discussed in LeSage and Pace (2009) and Golgher and Voss (2016).

interactions between $\hat{\beta}, \hat{\rho},$ and $\hat{\theta}$ simultaneously. The total effect, that which is observable, is the summation of both the direct and indirect effects. Take for example the parameter having at least 90 percent of income coming from self-employment activity (row 8 of Table 6.3), if we were to completely isolate each area, and thus cut off any influence between areas, we would need only rely on $\hat{\beta}$ derived from the OLS model and expect to see a 0.0422 standard deviation improvement in the percent of taxpayer

Table 6.4: Spatial Direct, Indirect, and Total Effects from the SDM for the Full Population [1][2]

Parameter [3]	Direct Effect	Indirect Effect	Total Effect
Making Retirement Contributions	0.7083	-0.4177	0.2906
Squared(Making Retirement Contributions)	-0.2950	0.5953	0.3003
Total Adjusted Gross Income	-0.1569	0.1560	-0.0010
Squared(Total Adjusted Gross Income)	0.0760	-0.2028	-0.1267
Qualifying for 20 Percent Credit	0.1075	0.4630	0.5705
Qualifying for 10 Percent Credit	0.4167	0.2046	0.6212
90 Percent or more of income from SE[4]	0.0345	0.3094	0.3439
Male	0.0531	0.1311	0.1842
Between 25 and 34	-0.0102	-0.1138	-0.1240
Between 35 and 44	-0.0281	0.0706	0.0425
Between 45 and 54	-0.0305	-0.0008	-0.0313
Between 55 and 59	-0.0327	0.1269	0.0942
Married filing Jointly	0.2309	0.2763	0.5072
Head of Household	0.1888	-0.1432	0.0456
1 to 3 dependents	-0.0467	0.1197	0.0730
4 or more dependents	0.0966	0.3031	0.3997
Using a paid preparer	-0.2502	1.0310	0.7808
Squared(Using a paid preparer)	0.3809	-1.1300	-0.7491

[1] Population reflects tax filers between 18 and 57 in 1999 with income below $100,000.
[2] Averaged across 2002-2013
[3] All parameters reflect the percent for a given ZIP Code except Total AGI, which is the sum of reported AGI
[4] SE - Self-Employed
Source: author's computation

participation with every one standard deviation increase in taxpayers earning a majority self-employment income. However, because ZIP Codes are not isolated and influence is allowed to flow freely across borders, the estimated $\hat{\beta}$ of 0.0293 for the SDM is only one parameter of three that define the observed relationship. The other two parameters, $\hat{\rho}$ being 0.4556 and $\hat{\theta}$ at 0.0462 are all included in the computation of direct and indirect effects, although here it is applied to a 7,435 × 7,435 weight matrix. Intuitively, the estimated direct effect from Table 6.4 estimates the expected marginal influence that the estimand has on participation from within ZIP Code *i*, similar to the interpretation of $\hat{\beta}$ in the OLS model. The estimated indirect effect from Table 6.4 is the expected marginal influence that a change in one estimand has on participation in all other ZIP Codes. In the case of having at least 90 percent of income coming from self-employment activity, the direct effect is 0.0345, the indirect effect is substantially larger at 0.3094, and the total effect is the summation of those, at 0.3439. That implies that the observed relationship is being more highly influenced by the impact of neighboring ZIP Codes than from within.

Of particular interest is the frequent incidence of opposing directions between the direct and indirect effects. Take for example, making a retirement contribution has a direct effect of 0.7083 with an indirect effect of -0.4277. This direct effect indicates that a one standard deviational increase in the percent of taxpayers making a retirement contribution within a ZIP Code is associated with a 0.7083 standard deviational increase in percent of Saver’s Credit participation within the same ZIP Code. However, the conflicting indirect effect indicates that an increase in the percent of taxpayers making a retirement contribution in a given ZIP Code negatively impacts neighboring locations' participation which simultaneously dampens participation in that initial ZIP Code. A different way to state this is to say those ZIP Codes with higher retirement contribution rates correspondingly have higher credit participation than those with lower contribution rates. This is indicated by the positive direct effect. But for ZIP Codes with neighbors having higher contribution rates than their own, they are more apt to have a lower prevalence of credit participation than their neighbors. This is seemingly contradictory given the correlation we have already established between retirement contribution behavior and Saver's Credit participation. To reconcile, while the true underlying mechanism driving this relationship is not clear, one assumption is that higher contribution areas are also higher income areas that find a larger portion of their population above the Saver's Credit income threshold and thus disqualified from participation. In this light, these findings become not just reasonable but expected.

When looking at the trends between direct and indirect effects, in most cases the indirect, or spillover, effects are a substantial portion of the total effect. Figure 6.2 maps the relationship between spillovers and total effects for each of the variables, presented in order of absolute total effect. Four parameters have negative spillovers; percent making retirement contributions, percent between 25 and 34, percent filing head of household, and percent between 45 and 54. Three parameters present spillovers that deflate the total effects on percent participating; percent making retirement contributions, percent filing head of household, and total AGI.

Focusing on the impact of marital status. Married filing jointly has a large effect relative to filing single or married filing separately. Head of household filing actually has a negative spillover effect on participation relative to single or married filing separately, which only serves to dampen, but not entirely eliminate the positive direct effects relative

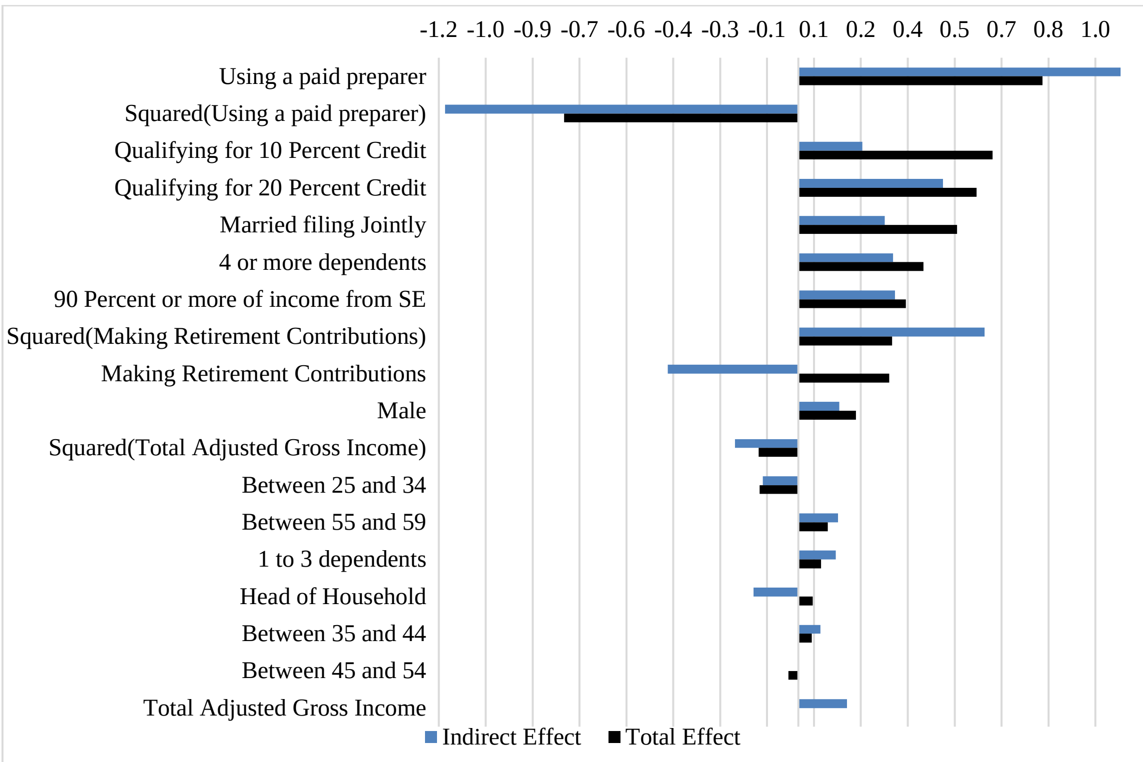


Figure 6.2: Distribution of Spatial Spillover Effects Relative to Total Effects for the SDM, Full Population

to filing single or married filing separately. Put in perspective, these indicate that ZIP Codes with more married filing jointly and head of household filers are associated with higher participation rates on average than single or married filing separately. ZIP Codes neighboring other ZIP Codes with more married filing jointly than single or filing separately have higher participation rates as well, but ZIP Codes neighboring areas with more head of household filers have lower participation rates. Here, again the mechanism for this result is unclear.

There are a few additional standout results worth noting. The percent using a paid preparer has both the largest total effect and the largest spillover effect, both positive. And, since the total effect is smaller than spillovers, the relationship is being entirely driven by spatial spillovers. The percent qualifying for 10 percent credit has the second largest total effect but represents the second smallest spatial spillover; two-thirds of the observed effect is being driven by direct effects, suggesting surprisingly little spatial interactions. Percent making retirement contributions overall is significant towards explaining participation, but spatial spillovers drastically undercuts what would otherwise be a very large effect.

These pooled results disguise some interesting annual trends, reported in Appendix D. Using a paid preparer initially has very large positive spillover effects. While spillovers remain positive over time, they generally have diminishing returns until 2009 when the trend shifts. Married filing jointly, relative to single or married filing separately, has consistently positive direct effects but has significantly larger spillover effects in the first year. These very large positive spillovers diminish quickly, and for a short time flip negative, but begin to trend up nearing the end of the study window. Keeping in mind that making a retirement contribution tends to be quite progressive, it is perhaps surprising that it initially presents with positive spatial spillovers. But, within two years spillovers have a generally negative impact on participation.

With remembering that this credit is specifically targeted at low- to middle-income households, it is more enlightening to examine the results of a targeted population set. I explore each subgroup in turn.

Section 6.3.2 - Subgroup 1

Recall, the Saver's Credit is, in application, more targeted than just at low- to middle- income households but also those presenting with a tax liability. Table 6.5 reports model estimates for this population of taxpayers. Implicit in parameter estimates shifting relative to the full population implicit is the dual impact of having income below the credit threshold as well as having a tax liability. Take for example the case of total AGI, we first observe a negative relationship between AGI and credit participation from Table 6.3. This is because as one's income surpasses the Saver's Credit threshold the ability to claim the credit disappears. However, when income is capped at the Saver's Credit income cutoff regardless of tax liability, estimates for the effect of total AGI is positive, suggesting as one's income increases to the limit the rate of participation increase.[23] For Subgroup 1 though, the estimated parameter is incorporating this positive income effect and the increased probability that the highest income individuals are more likely to also have a tax liability. So, when we explicitly control for having said liability, we see that within that group, the lower one's income is, the more likely she is to participate.

[23] The OLS estimates the coefficient at 0.0324 (statistically significant at the 0.001 level), SAR estimates it at 0.0128 (statistically significant at 0.05), SEM estimates 0.0139 (not statistically significant, and SDM estimates 0.0136 (not statistically significant).

As before, the OLS model suffers from assumption violations and can be rejected as an appropriate model specification for the underlying data. So too can the SAR model for the same reasons. In both cases there remains significant correlation in the model

Table 6.5: Subgroup 1 Model Estimates [1]

Parameter [2][3]	OLS	SAR	SER	SDM
Intercept	-0.0505	0.0021	0.0127	0.0033
	(0.0297) *	(0.0065)	(0.0294)	(0.0064)
Making Retirement Contributions	0.8041	0.7163	0.7351	0.7272
	(0.0296) ***	(0.0290) ***	(0.0292) ***	(0.0293) ***
Squared(Making Retirement Contributions)	-0.0844	-0.0557	-0.0548	-0.0506
	(0.0290) *	(0.0281) *	(0.0282) .	(0.0283) .
Total Adjusted Gross Income	-0.0048	-0.0161	-0.0056	-0.0054
	(0.0179) *	(0.0173) .	(0.0180)	(0.0181)
Squared(Total Adjusted Gross Income)	-0.0110	-0.0033	-0.0104	-0.0098
	(0.0167) .	(0.0162)	(0.0164)	(0.0164)
Qualifying for 20 Percent Credit	-0.0063	-0.0060	-0.0038	-0.0031
	(0.0072) .	(0.0070)	(0.0070)	(0.0070)
Qualifying for 10 Percent Credit	-0.0114	0.0039	0.0121	0.0173
	(0.0087) .	(0.0084) .	(0.0092) .	(0.0094) .
90 percent or more of income from SE [4]	-0.0586	-0.0377	-0.0492	-0.0454
	(0.0077) ***	(0.0076) **	(0.0082) **	(0.0084) **
Male	0.0220	0.0202	0.0179	0.0169
	(0.0076) **	(0.0074) *	(0.0075) .	(0.0075) .
Between 25 and 34	0.0576	0.0510	0.0560	0.0561
	(0.0130) **	(0.0126) **	(0.0129) **	(0.0130) **
Between 35 and 44	0.0646	0.0548	0.0539	0.0525
	(0.0096) **	(0.0093) **	(0.0095) **	(0.0096) **
Between 45 and 54	0.0521	0.0404	0.0367	0.0343
	(0.0109) ***	(0.0106) **	(0.0108) **	(0.0109) **
Between 55 and 59	0.0094	0.0043	0.0021	0.0002
	(0.0102) **	(0.0099) *	(0.0099) *	(0.0099) *
Married filing Jointly	0.4158	0.3631	0.3462	0.3363
	(0.0203) ***	(0.0198) ***	(0.0206) ***	(0.0207) ***
Head of Household	-0.0402	-0.0196	-0.0260	-0.0213
	(0.0245) *	(0.0238) *	(0.0249) .	(0.0251) .
1 to 3 dependents	0.0230	0.0332	0.0322	0.0323
	(0.0113) *	(0.0110) *	(0.0113) *	(0.0113) *
4 or more dependents	0.0087	0.0226	0.0031	0.0007
	(0.0173) .	(0.0168) .	(0.0175) .	(0.0176) .
Using a paid preparer	0.2591	-0.0451	-0.0156	-0.0031
	(0.2221) .	(0.0458) .	(0.0533) .	(0.0547) .
Squared(Using a paid preparer)	-0.2461	0.0687	0.1034	0.0981
	(0.2387) .	(0.0457) .	(0.0516) .	(0.0527) .
ρ		0.3111		0.6232
		(0.0153) ***		(0.0350) ***
λ			0.7568	
			(0.0264) ***	

Footnotes at end of table.

Table 6.5: Subgroup 1 Model Estimates, cont.

	OLS	SAR	SER	SDM
Lag(Making Retirement Contributions)				-0.2673
				(0.1610) .
Lag(Squared(Making Retirement Contributions))				-0.0571
				(0.1595)
Lag(Total Adjusted Gross Income)				-0.0302
				(0.0824)
Lag(Squared(Total Adjusted Gross Income))				0.0345
				(0.0892)
Lag(Qualifying for 20 Percent Credit)				-0.0030
				(0.0558)
Lag(Qualifying for 10 Percent Credit)				-0.0545
				(0.0365) .
Lag(90 Percent or more of income from SE[4])				0.0474
				(0.0264) .
Lag(Male)				-0.0165
				(0.0397)
Lag(Between 25 and 34)				-0.0600
				(0.0720)
Lag(Between 35 and 44)				-0.0222
				(0.0493)
Lag(Between 45 and 54)				-0.0019
				(0.0621)
Lag(Between 55 and 59)				-0.0237
				(0.0665)
Lag(Married filing Jointly)				-0.1047
				(0.0906)
Lag(Head of Household)				-0.0017
				(0.1054)
Lag(1 to 3 dependents)				-0.0574
				(0.0524)
Lag(4 or more dependents)				0.0216
				(0.0756)
Lag(Using a paid preparer)				0.0738
				(0.1304)
Lag(Squared(Using a paid preparer))				-0.1631
				(0.1304)
N	89,220	89,220	89,220	89,220
AIC	12,929	12,517	12,396	12,365
Moran I Residuals	0.0751 ***	0.0275 ***	(0.0030)	(0.0010)
Log Likelihood		-6,237.56	-6,175.73	-6,143.31
Lagrange Multiplier				
Error	1,515.30 ***			
LMlag	638.59 ***			
RLMerr	1,074.54 ***			
RLMlag	197.83 ***			

Standard Errors are in paretheses below parameter estimates
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Standard Errors in Paretheses below parameter estimates
[1] Averaged across 2002-2013
[2] All parameters reflect the percent for a given ZIP Code except Total AGI, which is the sum of reported AGI
[3] Lagged parameters are represented by θ in the SDM spatial model.
[4] SE - Self-Employed
Source: author's computation

residuals. Much as before, the SER and SMD models present with similar diagnostics suggesting that either would be more suited to explain the relationships than would OLS or SAR. Proper selection between the SEM and SDM though is less apparent with both presenting no autocorrelation in the residuals and similar AIC and Log Likelihood estimates. Also consistent are their parameter estimates.

However, it is important to keep in mind the model's spatial specification. The SER model allows for no feedback and assumes all observed patterns are expressly direct effects with no spillover. It is unreasonable, in most cases, to assume all spatial autocorrelation is contained entirely within the error. Here, this assumption can be rejected simply by examining the spatially lag parameter estimates derived from the SDM model. The lag on making a retirement contribution, qualifying for the 10 percent credit rate, and self-employed all present as statistically significant. As such, I focus the remaining analysis on the SDM.

Estimates of parameter affects are presented in Table 6.6. The total effect from retirement contributions is by far the largest influence on participation rates. The next largest effects come from married filing jointly, using a paid preparer, and qualifying for the 10 percent credit rate. The square parameter for making retirement contribution and using a paid preparer are both among the largest effects but are in the opposite direction of their linear counterparts. The interpretation of these squared parameters is similar to that of a more traditional model polynomial coefficient. So, in this case of making retirement contributions, generally there is a positive effect on participation but at a decreasing rate. Using a paid preparer, similarly, is associated with positive returns on participation but at a decreasing rate.

With respect to the impact of spatial spillovers, the ordering remains generally the same as that of total effects. The top six parameters for total effects are also the top six parameters for indirect effects but the ordering of magnitudes shift slightly with the square of both using a paid preparer and making a retirement contribution moving up.

Table 6.6: Spatial Direct, Indirect, and Total Effects from the SDM for Subgroup 1 [1]

Parameter [2]	Direct Effect	Indirect Effect	Total Effect
Making Retirement Contributions	0.7309	0.5715	1.3023
Squared(Making Retirement Contributions)	-0.0525	-0.2965	-0.3490
Total Adjusted Gross Income	-0.0059	-0.0737	-0.0796
Squared(Total Adjusted Gross Income)	-0.0094	0.0578	0.0484
Qualifying for 20 Percent Credit	-0.0032	-0.0244	-0.0276
Qualifying for 10 Percent Credit	0.0165	-0.1352	-0.1187
90 Percent or more of income from SE[3]	-0.0450	0.0570	0.0120
Male	0.0168	-0.0131	0.0037
Between 25 and 34	0.0556	-0.0735	-0.0179
Between 35 and 44	0.0526	0.0134	0.0659
Between 45 and 54	0.0344	0.0289	0.0633
Between 55 and 59	-0.0002	-0.0776	-0.0778
Married filing Jointly	0.3381	0.2962	0.6343
Head of Household	-0.0216	-0.0411	-0.0627
1 to 3 dependents	0.0317	-0.1009	-0.0692
4 or more dependents	0.0011	0.0617	0.0628
Using a paid preparer	-0.0017	0.2188	0.2171
Squared(Using a paid preparer)	0.0962	-0.3003	-0.2041

[1] Averaged across 2002-2013

[2] All parameters reflect the percent for a given ZIP Code except Total AGI, which is the sum of reported AGI

[3] SE - Self-Employed

Source: author's computation

Though, if we ignore those parameters, the ordering of the remaining four is entirely the same; the largest spillover effects are observed for the percent making retirement contributions, followed by percent married filing jointly, percent using a paid preparer, and finally percent qualifying for the 10 percent rate relative to qualifying for the 50 percent rate.

The spillover effects for these six parameters are also in the same direction as total effects indicating that spatial spillovers work to amplify, or entirely drive, the total effects we observe. This is clearly represented in Figure 6.3. Only in two parameters, using a paid preparer and qualifying for the 10 percent credit rate relative to the 50 percent rate, do the direct and indirect effects work in conflict. However, using a paid preparer's negative direct effect is so small as to be insignificant in affecting the total effect, at -0.0017. Qualifying for the 10 percent credit's direct effect is also much smaller than its spillover counterpart, at 0.0165 and -0.1352 respectively.

Figure 6.3 also serves to highlight the relatively inconsequential economic impact that remaining twelve parameters provide. Of the eighteen parameters estimated, only the top six influence the percent of participation more than 0.1 standard deviations in either direction. This implies that other observable demographics, including age, gender, and number of dependents, have little impact on participation. Also of little economic impact is the total size of AGI with a small negative effect on participation both directly and indirectly. Given that this population has been limited to only those with income below the Saver's Credit income qualifying threshold and those with a tax liability, this suggests that qualifying for the 50 percent credit is a large driver in participation. In Chapter 7, I explore further the impact of each credit rate on behavior and confirm that the 50 percent rate has the largest influence on behavior.

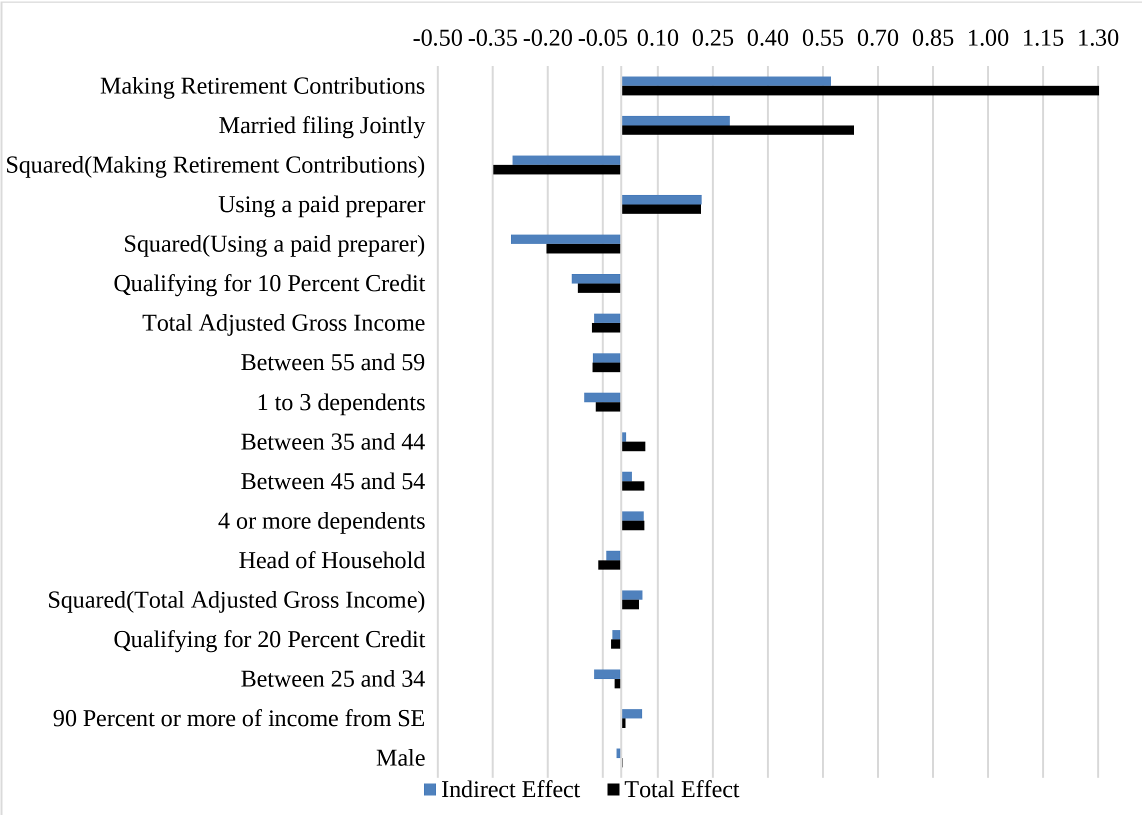


Figure 6.3: Distribution of Spatial Spillover Effects Relative to Total Effects from the SDM, Subgroup 1

The pooled estimates reported in Table 6.6 and Figure 6.3 obfuscate year over year trends that prove to be quite illuminating. Ignoring the quadratic estimands, Figure 6.4 shows the annual direct, indirect, and total effects of the remaining four largest parameters. In each case we see large swings in the indirect effect, represented by solid black lines, but a relatively stable pattern in the direct effects, represented by dotted black lines. Further, the total effects all appear to be heavily influenced by indirect effects regardless of year. In three of the four instances, spillovers present high in the initial year and generally taper off over time. Only the percent qualifying for the 10 percent credit, (box d), show the reverse. While speculative, this is probably the result of the expansion in awareness of the credit as time passes. What all of this uncovers is that spillovers play a significant role in the observed total effect of participation and these parameters in particular drive much of the patterns of participation that we see.

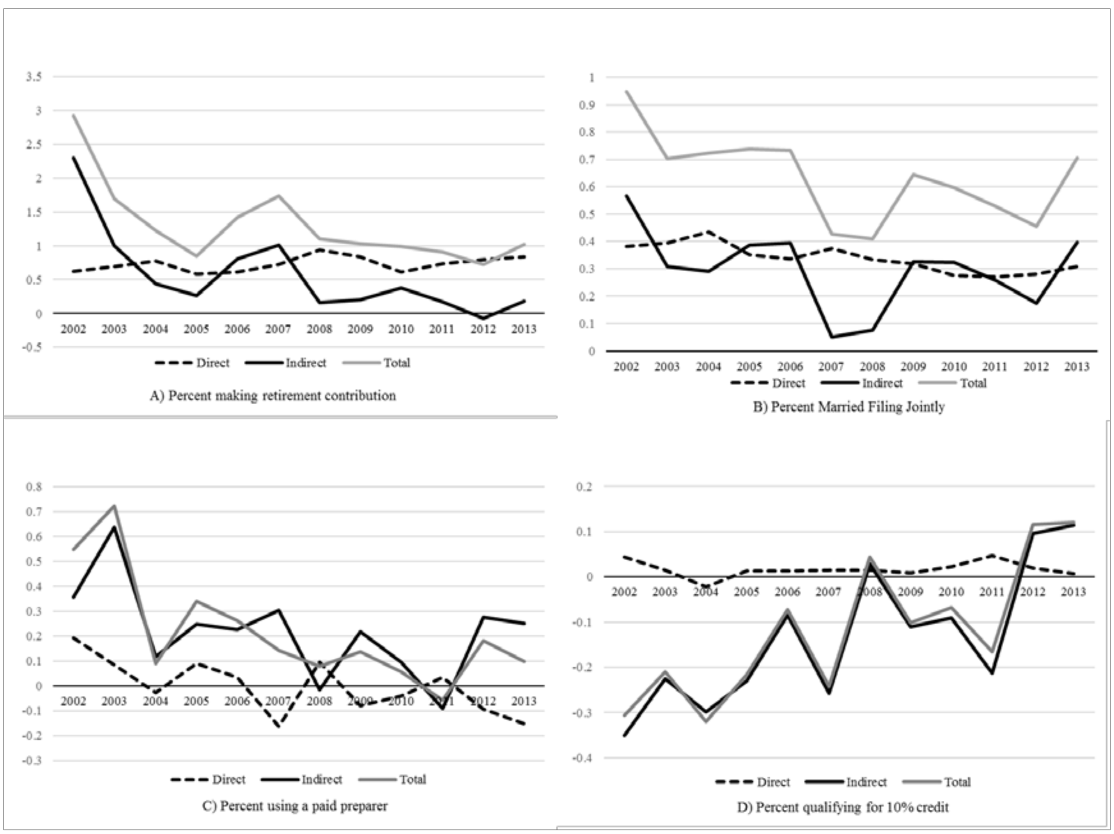


Figure 6.4: Annual Effects from the SDM, Subgroup 1, 2002-2013

Section 6.3.3 - Subgroup 2

The last population of interest are those meeting all criteria for claiming the credit; income is below the threshold, presence of a tax liability, and making a retirement contribution. Since, by definition, all ZIP Codes already have 100 percent retirement contribution participation, this parameter has been removed from the model. Results from the four models are presented in Table 6.7. And again, model diagnostics suggest the SDM remains the best fit.

Table 6.7: Subgroup 2 Model Estimates [1]

Parameter [2][3]	**OLS**	**SAR**	**SER**	**SDM**
Intercept	0.0000	0.0041	0.0162	0.0055
	(0.0107)	(0.0104)	(0.0385)	(0.0104)
Total Adjusted Gross Income	0.0456	0.0436	0.0289	0.0305
	(0.0238) *	(0.0230) .	(0.0243) .	(0.0245) .
Squared(Total Adjusted Gross Income)	-0.0373	-0.0413	-0.0317	-0.0329
	(0.0227) .	(0.0220) .	(0.0224) .	(0.0225) .
Qualifying for 20 Percent Credit	0.0156	0.0156	0.0132	0.0133
	(0.0118) .	(0.0114) .	(0.0114)	(0.0114) .
Qualifying for 10 Percent Credit	0.0267	0.0442	0.0359	0.0390
	(0.0127) *	(0.0123) *	(0.0128) .	(0.0129) *
90 percent or more of income from SE [4]	-0.0286	-0.0032	-0.0116	-0.0080
	(0.0114) *	(0.0111) .	(0.0113) .	(0.0113) .
Male	0.0463	0.0445	0.0547	0.0550
	(0.0122) **	(0.0118) **	(0.0120) **	(0.0120) **
Between 25 and 34	0.1606	0.1376	0.1508	0.1480
	(0.0184) **	(0.0179) **	(0.0181) **	(0.0181) **
Between 35 and 44	0.1510	0.1251	0.1331	0.1285
	(0.0170) **	(0.0165) **	(0.0167) **	(0.0167) **
Between 45 and 54	0.1203	0.0941	0.0987	0.0947
	(0.0176) **	(0.0170) **	(0.0172) **	(0.0172) **
Between 55 and 59	0.0785	0.0654	0.0674	0.0638
	(0.0152) **	(0.0147) *	(0.0147) *	(0.0148) *

Footnotes at end of table.

Table 6.7: Subgroup 2Model Estimates, cont.

	OLS	SAR	SER	SDM
Married filing Jointly	0.3089	0.2451	0.2429	0.2366
	(0.0238) ***	(0.0232) ***	(0.0238) ***	(0.0239) ***
Head of Household	0.0778	0.1217	0.1142	0.1289
	(0.0282) *	(0.0273) **	(0.0282) *	(0.0284) **
1 to 3 dependents	0.0441	0.0449	0.0534	0.0543
	(0.0158) **	(0.0153) *	(0.0154) *	(0.0155) *
4 or more dependents	-0.0175	-0.0005	0.0078	0.0115
	(0.0205) .	(0.0198) .	(0.0206) .	(0.0207) .
Using a paid preparer	0.2302	0.1921	0.2066	0.1956
	(0.0495) **	(0.0479) *	(0.0503) *	(0.0507) *
Squared(Using a paid preparer)	-0.1064	-0.0994	-0.0698	-0.0617
	(0.0495) *	(0.0479) .	(0.0496) .	(0.0498) .
ρ		0.6026		0.5655
		(0.0277) ***		(0.0383) ***
λ			0.7114	
			(0.0297) ***	
Lag(Total Adjusted Gross Income)				0.0448
				(0.1062)
Lag(Squared(Total Adjusted Gross Income))				-0.0270
				(0.1217)
Lag(Qualifying for 20 Percent Credit)				0.0128
				(0.0942)
Lag(Qualifying for 10 Percent Credit)				-0.0110
				(0.0614)
Lag(90 Percent or more of income from SE[4])				-0.0082
				(0.0637)
Lag(Male)				-0.1053
				(0.0598) .
Lag(Between 25 and 34)				-0.1526
				(0.1189) .
Lag(Between 35 and 44)				-0.1410
				(0.1075)
Lag(Between 45 and 54)				-0.0205
				(0.1177)
Lag(Between 55 and 59)				-0.1117
				(0.1156)
Lag(Married filing Jointly)				0.2121
				(0.1334) .
Lag(Head of Household)				0.1208
				(0.1554)
Lag(1 to 3 dependents)				-0.0900
				(0.1012)
Lag(4 or more dependents)				-0.0937
				(0.1226)
Lag(Using a paid preparer)				0.0183
				(0.1993)
Lag(Squared(Using a paid preparer))				-0.1023
				(0.2016)
N	7,435	7,435	7,435	7,435
AIC	18,460	19,532	19,513	19,482
Moran I Residuals	0.0680 ***	0.0027 *	-0.0038	-0.0021
Log Likelihood		-9,747	-9,738	-9,706
Lagrange Multiplier				
Error	1,293 ***			
LMlag	1,029 ***			
RLMerr	361 ***			
RLMlag	96 ***			

Standard Errors are in paretheses below parameter estimates
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Standard Errors in Paretheses below parameter estimates
[1] Averaged across 2002-2013
[2] All parameters reflect the percent for a given ZIP Code except Total AGI, which is the sum of reported AGI
[3] Lagged parameters are represented by θ in the SDM spatial model.
[4] SE - Self-Employed
Source: author's computation

Moving directly into parameter effects, Table 6.8 shows the direct, indirect, and total effects for the SDM model. When we look across the variables, indirect effects are again largely driving the total effects observed. The two largest contributors are married filing jointly and filing head of household. To put this into context, these filing statuses are in relation to filing single or married filing separately. This implies that of the three filing statuses, after controlling for all factors that would qualify someone for Saver's Credit participation, married filing jointly and filing head of household strongly improve the expectation of credit participation.

When comparing to the previous Subgroup, we observe some drastic shifts in parameter importance. The percent qualifying for the 10 percent credit was one of the top estimands for Subgroup 1, but here it falls entirely out of relevance. Filing head of household was previously small and negatively related relative to other effects, but now represents the second largest effect, indirect or otherwise. Another major shift is the growing importance of total AGI; previously total AGI was marginally and negatively related to credit participation, but now it is among the top five effects, and positive. Percent with four or more dependents now appears to have a much larger, negative, affect. And percent male was previously entirely economically irrelevant to participation, but now reports some (negative) influence.

Many of the parameters explored are reported relative to some other (omitted) parameter. When we look across these groups, interesting patterns emerges. As has already been discussed, married filing jointly and filing head of household are both relative to single or married filing separately and report all positive impacts on

Table 6.8: Spatial Direct, Indirect, and Total Effects from the SDM for Subgroup 2 [1]

Parameter [2]	Direct Effect	Indirect Effect	Total Effect
Total Adjusted Gross Income	0.0318	0.1988	0.2305
Squared(Total Adjusted Gross Income)	-0.0338	-0.1473	-0.1811
Qualifying for 20 Percent Credit	0.0133	0.0157	0.0289
Qualifying for 10 Percent Credit	0.0390	0.0054	0.0443
90 Percent or more of income from SE[3]	-0.0081	-0.0196	-0.0277
Male	0.0538	-0.1899	-0.1360
Between 25 and 34	0.1463	-0.2583	-0.1120
Between 35 and 44	0.1271	-0.2236	-0.0965
Between 45 and 54	0.0945	-0.0096	0.0849
Between 55 and 59	0.0620	-0.2773	-0.2153
Married filing Jointly	0.2421	0.9027	1.1448
Head of Household	0.1321	0.5360	0.6681
1 to 3 dependents	0.0534	-0.1578	-0.1044
4 or more dependents	0.0101	-0.2373	-0.2272
Using a paid preparer	0.1976	0.3365	0.5341
Squared(Using a paid preparer)	-0.0639	-0.3600	-0.4239

[1] Averaged across 2002-2013

[2] All parameters reflect the percent for a given ZIP Code except Total AGI, which is the sum of reported AGI

[3] SE - Self-Employed

Source: author's computation

participation. The number of dependents, 1 through 3 and 4 or more, are relative to having no dependents. Here, the direct effect estimate suggests that ZIP Codes with more reported dependents have a greater prevalence of credit participation but ZIP Codes with neighbors having more reported dependents have a lower prevalence of participation. Looking at the distribution of age, we again observe this pattern of conflicting direct and indirect effects. For all age groups, relative to extremely young and old, there is a positive direct effect with negative indirect effects that drive total effects to be negative. The

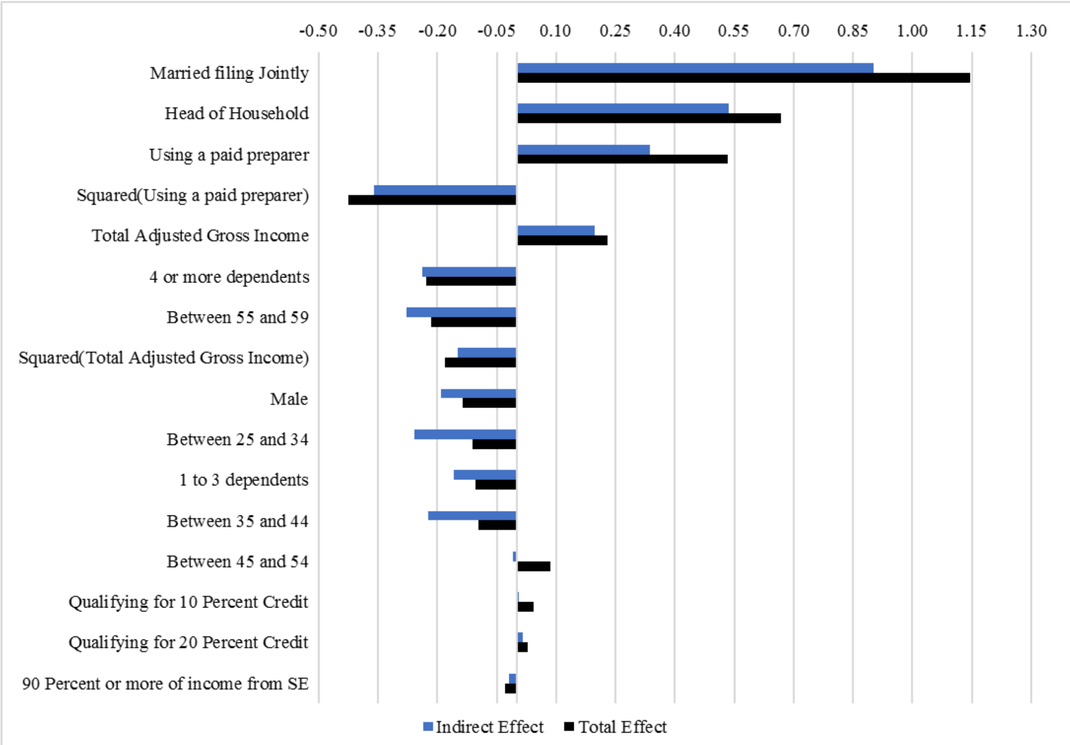


Figure 6.5: Distribution of Spatial Spillover Effects relative to Total Effects from the SDM, Subgroup 2

intuition behind these dependent and age demographic findings is not entirely clear and begs further investigation. Lastly, relative to qualifying for the 50 percent credit rate, qualifying for the 10 percent or 20 percent rates are both positively associated with higher credit participation. This also aligns with the positive estimate for total AGI.

Two parameters, while retaining their directional effect, exhibit substantial growth in intensity; married filing jointly and using a paid preparer. The latter's total effect grew by nearly 150 percent from its previous estimate. Married filing jointly grew by approximately 80 percent. Married filing jointly's effect now increases participation by more than 1 standard deviation per one standard deviation improvement in percent filing married filing jointly.

Figure 6.5 graphically presents the indirect and total effects of the SDM model for Subgroup 2. This reiterates and amplifies the earlier statement of spillovers driving overall effects. But for two, there are no parameters that exhibit less than 50 percent of the total effect coming from spillovers. Moreover, for six parameters spillovers drive entirely the observed total effect. Another important observation is the expansion in the number of parameters that have a substantial impact on participation. Where before most parameters hovered around +/- 0.05 standard deviations, now all but two surpass that threshold.

As before, these pooled results hide underlying variability across time. Figure 6.6 shows the four largest parameter effects broken out by year. Across all variables, direct effects are quite stable over time and offer little in the way of explaining total effect. Spillover effects are the main drivers in total effects for every year of this study. We also see that the initial years of the Saver’s Credit availability are the years in which most spatial spillovers occurred and these indirect effects generally taper off over time.

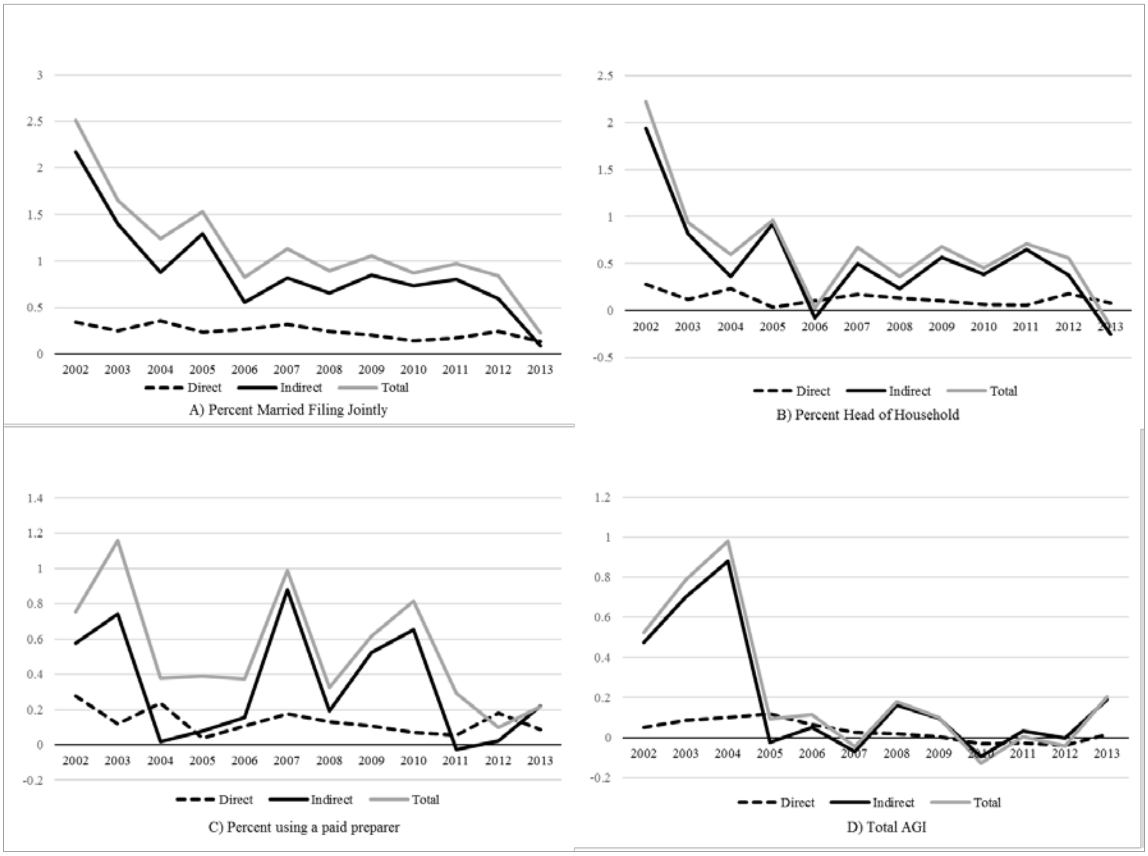


Figure 6.6: Annual Effects from the SDM, Subgroup 2, 2002-2013

Section 6.4 - Discussion

In this study, I explicitly examine the spatial influence that filing characteristics have on Saver's Credit participation. In the earlier studies, I was able to establish the existence and isolate the locations of spatial autocorrelation. In this study, I honed in on filing characteristics associated with participation and quantified the direction and magnitude of the spatial spillovers derived from those characteristics. I compared four models, one that assumes independence and zero autocorrelation, and three that relax these assumptions. I prove through derivation and application that the more traditional OLS model assumptions of independence and spherical errors is inappropriate for highly spatially granular data and leads to, at a minimum, erroneous standard errors and, more often, omitted variable biases. To correct for these violations, I employed three different spatial regression models, the SAR model, the SEM model, and the SDM model.

The advantage of spatial regression models is their ability to control for and quantify spatial correlations embedded in the error term. These spatial relationships are ignored in the OLS model. As I found, for each of my four population definitions the OLS model was routinely rejected for having spatial autocorrelation in the residuals. The spatial model that most closely fit the data was consistently the SDM. In this model, spatial lags are accounted for in the dependent variable as well as all explanatory variables.

Previous investigations into the Saver's Credit all but ignored the factors contributing to participation. Ramnath (2013), within a larger context of examining the effect of the credit on retirement contributions, was the only one to explore the drivers to participation but never explored the spatial aspect influencing participation. No other study has directly explored the spillover effects that location has on federal tax policy, let alone Saver's Credit participation. Surprisingly little work has taken up the task of exploring the interaction between federal tax policy and geography. The few notable exceptions focus on the spatial impact of either tax evasion or tax burden, or how federal tax effects localities. This marks the first study to quantify the impact that spatial spillovers exert on federal tax policy participation.

My results consistently found that spatial spillovers are a significant contributor towards participation, regardless of the definition of eligibility. This is true particularly for Subgroup 2 where spillovers represented nearly all observed total effects. When distributed by year, spillovers' impact diminished over time as well. This is consistent with findings from the previous sections suggesting that as time passes, clustering of activity also dissipates.

Where spillovers were consistently uncovered, the exact parameters most affecting participation also remained consistent across the various eligibility groups. The full population was most affected by using a paid preparer and this was driven entirely by spillovers, with spillovers representing more than a one standard deviation improvement on participation given a one standard deviation increase in the parameter. Subgroup 1 saw the top impact replaced by positive effects from retirement contributions which were more driven by direct effects than spillovers. Subgroup 2 was most impacted by married filing jointly, showing strong positive total and spillover effects.

The parameters persistently showing the largest spillover effects included percent reporting married filing jointly, percent making retirement contributions, and percent using a paid preparer. The latter two had a nonlinear relationship with participation as represented by opposite influence from their squared counterparts. The remaining parameters varied in the size of their impact but generally remained less significant, and in some instances entirely economically insignificant.

When divided by year, the first few years of credit availability routinely showed the largest spatial effects, positive or negative, that subsequently diminished over time. In most cases these large spatial effects also represent the majority of total effects. Witnessing these time trends serves to validate the assumptions made in earlier studies that as time passes, and more individuals become aware of the Saver's Credit, the effects of location decline in importance.

This paper remains empirically silent on the exact mechanisms explaining this serial decline. One explanation could be the close clustering of financially literate populations, or utilized paid preparers, being aware of and responding to incentives offered by the introduction of credit and as knowledge expanded the impact of location became less significant. Another could be a close clustering of populations that were already choosing to making retirement contributions and saw this credit as a way to lower tax liability without changing behavior. Answering this question is taken up in Chapter 7 below.

Here, I measured the global spatial effect. One could deconstruct these global effects through a Graphically Weighted Regression and thus estimate local spatial spillover effects. It is unclear how impactful such an exercise would be. Taking into account the results from Chapter 5, it is reasonable to assume that areas reporting high autocorrelation will also report high local spatial effects and areas with low autocorrelation would report smaller spatial effects. Perhaps the added value of disaggregating the global spatial effect would be to quantify the size of the spillover.

The findings from this paper significantly contribute to our understanding of how regions respond to tax expenditures. Previous research ignored the influence of spatial interactions as a primary contributor toward observed tax expenditure take up. Given the expanse in high quality, spatially granular data that is now available, it is critical that these spatial relationships not be overlooked. I highlight the error in doing so both mathematically and empirically and directly measure the direction and magnitude of spatial influence across borders. Using these findings, we now know that participation in the Saver's Credit is not only directly related to neighboring participation but also related to the activity and characteristics of those neighbors. Left still to understand is the underlying mechanism driving these correlations and whether the credit has induced any new retirement contributions. In the next section I take on the latter of these two remaining questions.

CHAPTER 7 - THE IMPACT ON RETIREMENT SAVINGS

The Saver's Credit was designed to influence retirement contributions, but up to this point I have remained silent on whether the credit has had the desired effect. I have shown there is certainly a correlation between retirement contributions and participation but nothing beyond this cursory correlation. In fact, it remains unclear whether those utilizing the credit were responding to the credit's incentives by increasing savings or if they were merely capitalizing on an opportunity to lower their tax liability after having already made contributions. The question left unresolved is whether the credit has actually created any new retirement savings.

This study sets out to answer that question. To do so, I apply a regression discontinuity approach by exploiting the three sharp income notches used in determining credit rate eligibility. A regression discontinuity (RD) model is ideal in this situation because taxpayers are deemed (in)eligible for the credit based on sharp notch-points along the income distribution. As such, this design tests the discontinuity of savings contributions at these particular notch-points to identify the causal effect of the Saver's Credit.

As I highlighted in Chapter 2, measuring the effects of the Saver's Credit is not a new endeavor. Duflo et al. (2007) found the credit averaged at best an $81 increase in contributions and at worst a $32 decline. Ramnath (2013), through also employing an RD model, estimated the impact of the credit to have at best improved contributions an average of $127 or at worst lowered contributions $356. Shortly thereafter, Heim and

Lurie (2014) estimated the Saver's Credit produced an overall 2.8 percentage point increase in the probability to contribute and an estimated increase of $317 of contributions.

In none of these studies were more than a handful of years allowed to pass before the credit's impact was assessed. As Chapters 4 and 5 document, the Saver's Credit has experienced a far more gradual expansion that make this an opportune moment to reevaluate the earlier results. These previous studies assumed a rather short time horizon of only 4 or 5 years for taxpayers to fully learn and respond to the Saver's Credit incentives. Figure 4.1 documents the gradual expansion of participation starting at 51 percent of filers fully eligible to take the credit doing so in the initial year, and reaching just above 60 percent by 2006, the year for which most of these previous investigations rely. Allowing another 7 years to pass, participation has improved by nearly 10 percentage points. In this research I reassess the impact of the Saver's Credit and overcome some of research's earlier limitations.

Further, Ramnath's study, for which this investigation emulates, suffers from the additional limitation of only being able to observe information reported on the tax return. This results in her only being able to impute retirement contributions based on observed Saver's Credit claiming activity and reported Traditional IRA contributions. My analysis relies on far more comprehensive retirement savings contribution information. There is evidence to suggest there may be between 12 to 13 times more individuals participating in 401(k)-type plans than Traditional IRA plans (Pierce and Gober 2013; Bryant and Gober 2013). By using information returns from the IRS, I will be able to examine IRA, 401(k), and other DC savings contribution activity to better assess the true effect of the Saver's Credit.

For this section of analysis, I adjust how I analyze the data in three substantive ways. Where before I focused on subdividing the population into Subgroups and exploring the various changes in participation through that lens, here I return to focus only on the full population of taxpayers. Where before I focused much attention on the Southeast region, here I drop that focus to include all taxpayers in a given tax year, regardless of filing location. Lastly, I normalize the data by computing the distance away from any notch-point. While the Saver's Credit has three distinct notch-points, relating to qualifying for a 50 percent credit rate, 20 percent rate, or 10 percent rate, I in effect aggregate over these three limits by measuring the distance away from the income cutoff. Throughout the analysis I do reference specific impacts for each of the credit rates, but for the crux of the analysis I focus on overall effects.

The benefit of an RD approach is its ability to derive impact from sharp discontinuities between income and credit qualification at discrete notch-points. As the name implies, this approach measures the discontinuity, here of retirement contributions, at and around the sharp credit qualification income cutoff. Figure 7.1A presents this distribution for 2000 and 2002 for the full population, using a $50 bin width[24]. It is obvious from this simple plot that there is clear discontinuity in 2002 at the notch-point.

---

[24] Each point along the distribution represents the average retirement contribution amount for taxpayers who have been grouped together in $50 AGI increments. The distance from threshold can be interpreted as $50 increments from the AGI threshold. For example, a reported distance of -1 includes all taxpayers with a difference in income between their reported AGI and the credit threshold equaling between -$1 and -$50.

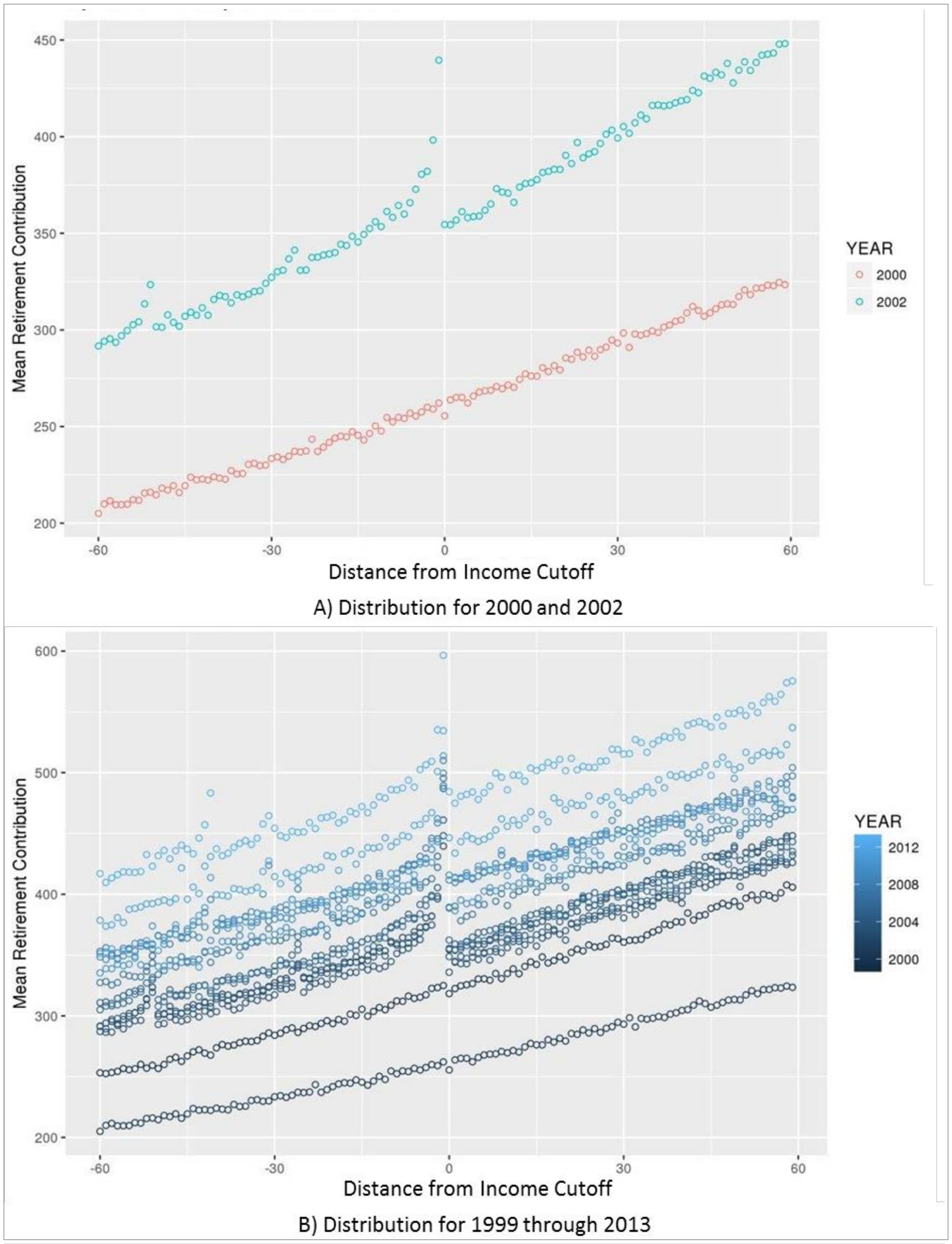


Figure 7.1: Distribution of Average Retirement Contributions around the Saver's Credit Income Cutoff by Year

Expanding this figure to include all tax years, Figure 7.1B, the pattern of discontinuity holds for all years after the introduction of the Saver's Credit.

I model this relationship with,

$$A_i = \alpha + \tau D_i + \beta_0(Y_i - c) + \beta_1(Y_i - c) * D_i + \varepsilon_i \qquad (7.1)$$

where A_i is the average retirement savings contributions for bin i, D_i is an indicator for being eligible to receive the credit with D = 1 when the filer is eligible and 0 otherwise, Y_i is average AGI, and c is the income threshold for receiving the credit. The primary treatment effect, τ, is

$$\tau = E[A_i | D = 1] - E[A_i | D = 0] \qquad (7.2)$$

In other words, the effect of the Saver's Credit is the difference between retirement savings contributions below the notch and contributions above.

While the RD model is well suited in principle, one critical assumption is that the density of the continuous variable, here AGI, is smooth across the notch-point. In other words, we should not observe a swell of filers just left of the notch-point or a dip in filers just right. Such findings would violate the assumption that filers below and above the notch are "as good as randomly assigned" (Lee and Card 2008) and therefore limit the ability to assign causal effect on retirement savings squarely on the Saver's Credit. This bunching could suggest a reverse causality bias; individuals who are contributing to retirement plans are incentivized to manipulate their reported income in order to qualify for the credit rather than the Saver's Credit inducing individuals to contribute to a retirement plan.

Because the income cutoffs for qualifying for the credit are known and publicly available prior to filing, and AGI is self-reported by the taxpayer with only limited third-party reporting, there is an apparent incentive to manipulate income to locate just below the notch-point. This affords the taxpayer the ability to both qualify for and perfectly optimize the benefits of the Saver's Credit. As I present in the proceeding section, there is significant evidence point to a violation of this assumption. In fact, I employ two quantitative measures that both confirm the presence of bunching.

I overcome this setback by bounding the treatment effect using a technique developed by Sallee (2011) and replicated in Ramnath (2013). To do this, I remove the taxpayers with the largest retirement contributions from the analysis, as they ostensibly have the most credit value to gain from locating below the income cutoff. This produces a lower bound estimate, or a floor for which the true treatment effect will not be below. I then remove the taxpayers with the smallest retirement contributions from the analysis, as they ostensibly represent taxpayers who have sacrificed the least in current income flows to qualify. This produces an upper bound estimate, or a ceiling for which the true treatment effect will not be above. A more thorough description is presented later.

After employing this approach, I find that the Saver's Credit has had an impact on retirement contributions. I estimate an average retirement contribution improvement of between $55 and $84 per person per year. When disaggregated by credit rate, I uncover that the 50 percent rate and the 10 percent rate are the drivers behind these findings with an average improvement ranging between $70.81 and $108.02, for the 50 percent rate, and $52.92 and $102.28 for the 10 percent rate. While these findings may appear inconsequential in absolute terms, they represent an approximate 12 to 24 percent improvement in the average size of retirement contributions from taxpayers just above the credit's AGI cutoff. These findings are evidence that the Saver's Credit is having a marginal impact on retirement contributions.

The remainder of this study proceeds as follows: Section 7.1 formally tests the RD assumption and quantifies the size of bunching. Section 7.2 presents empirical RD results employing a number of different methods to adjustment for the bunching behavior. And finally, Section 7.3 provides a discussion.

Section 7.1 – Assessing Bunching

When using an RD approach, it is imperative that the population density around the income notch-point be smooth, otherwise it is unclear whether the credit is inducing more retirement savings contributions or if contributions are causing manipulation of income in order to qualify for the Saver's Credit. Figure 7.2A presents the density plot of the number of taxpayers around the AGI notch-point for years 2000 and 2002. Each point represents the total number of taxpayers with a bin width of $50. It is quite apparent that the running variable is discontinuous in 2002. When expanding to the full set of tax years post credit introduction, Figure 7.2B, this discontinuity is maintained for most years. Only tax years 2009-2013 appear to deviate from this obvious discontinuity pattern. Although, as we will see, there still remains measurable discontinuity in these years.

Given these visual cues, it is reasonable to assume bunching, or more precisely a manipulation of AGI, is occurring. Without addressing this violation to the RD identification condition, the results obtained from merely applying the derived model will be biased for finding an effect when in reality the causal relationship is reversed. However convincing this visual approach may be, it only suggests the presence of

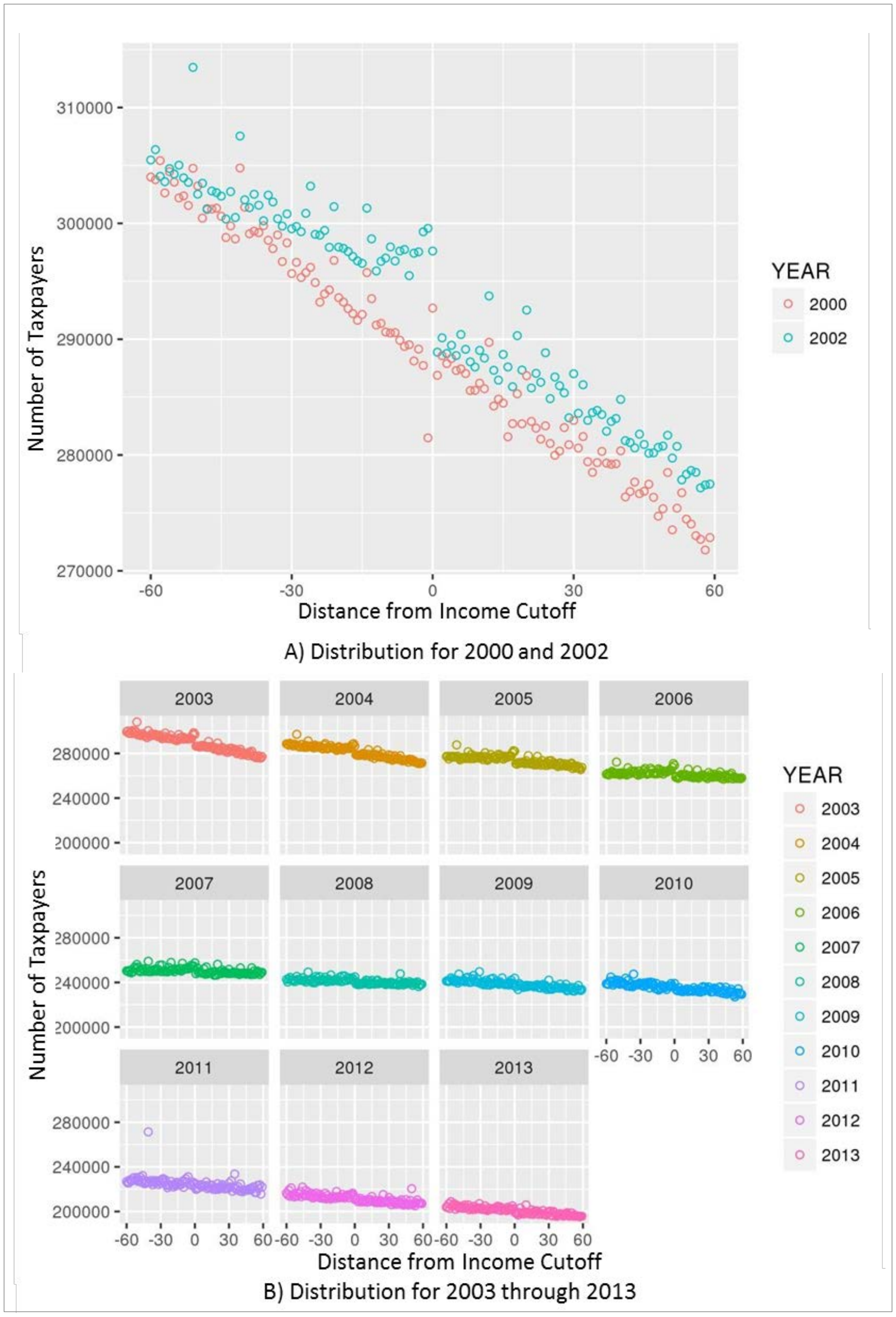


Figure 7.2: Density Distribution for the Number of Taxpayers around the Saver's Credit Income Cutoff by Year

bunching, it does not quantify the size nor definitively conclude bunching for years that appear more ambiguous, such as 2009-2013.

The next two subsections take a more quantitative approach. The first measures the discontinuity in the distribution of the running variable across the notch-point. The next estimates the magnitude of bunching.

Section 7.1.1 – Regression Density Discontinuity

In this approach to estimating the presence of discontinuity, I apply the normal Regression Discontinuity design to the density distribution. This approach mimics the McCrary test (2008). I run a local linear regression taking the form:

$$F_i = \alpha + \tau D_i + \beta_0(Y_i - c) + \beta_1(Y_i - c) * D_i + \varepsilon_i \qquad (7.3)$$

where F_i is the number of taxpayers in bin i, D_i is the indicator of location, either below or above the notch-point, with D = 1 for below and 0 otherwise. Notice, this model is the same as equation (7.1), with only dependent variable altered. Thus, the estimate of interest is again τ, but here τ is

$$\tau = E[F_i|D = 1] - E[F_i|D = 0] \qquad (7.4)$$

I rely on the method developed in Imbens and Kalyanaraman (2009) to derive the optimal bandwidth for smoothing.[25] To check for robustness, I also use half and double the estimated optimal bandwidth. Table 7.1 presents the estimated density discontinuity for Tax Years 1999 through 2013 for all three bandwidth estimates. Values greater than

[25] As adapted from Imbens and Kalyanaraman (2009), the optimal bandwidth, *h*, is equal to

$$\hat{h}_{opt} = C_K * \left(\frac{2\hat{\sigma}^2(c)}{\hat{f}(c) * \left(\left(\widehat{m}_+^{(2)}(c) - \widehat{m}_-^{(2)}(c) \right)^2 + (\hat{r}_+ + \hat{r}_-) \right)} \right)^{1/5} * N^{-1/5}$$

where $\hat{h}_{opt}$ is the optimal bandwidth, C_K is a constant that depends on the kernel, *K*, used, $\hat{\sigma}^2(c)$ is the estimated conditional variance at the threshold, $\hat{f}(c)$ is the estimated density at the threshold, $\widehat{m}_+^{(2)}(c)$ and $\widehat{m}_-^{(2)}(c)$ are the limits of the second derivatives at the threshold from the left and right, $\hat{r}_+$and $\hat{r}_-$ are regularization terms, and *N* is the number of taxpayers.

Table 7.1: Estimated Number of Excess Taxpayers Below the Income Cutoff, 1999-2013

	Size of Bandwidth [1]					
Year	Optimal		Half Optimal		Double Optimal	
1999	-1,345		-3,784		-453	
2000	-2,891	*	-4,909	*	-1,618	*
2002	6,595	***	6,609	**	5,362	***
2003	5,820	**	6,321	*	4,469	***
2004	5,959	***	6,733	**	4,576	***
2005	5,474	**	6,164	*	4,242	***
2006	4,908	**	6,113	*	4,059	***
2007	3,097	*	2,662		3,035	**
2008	3,082	***	3,087	*	2,919	***
2009	909		-160		1,043	
2010	1,875	.	2,177		1,532	*
2011	2,955	.	1,215		2,919	*
2012	4,213	***	4,783	***	2,841	***
2013	3,497	***	4,685	***	2,378	***

[1] Bandwidth is determined by employing Imbens-Kalyanamaran algorithm.
Signif. Codes: 0 '***' 0.001 '**', 0.01 '*', 0.05 '.' , 0.1 ' ', 1
Source: author's computation, Compliance Datawarehouse, IRS

zero represent the estimated additional number of taxpayers above what would be expected if the density of reported AGI was continuous across the credit qualifying notch-point. These positive estimates indicate bunching and thus violating the RD assumption of a smooth running variable. Negative values indicate fewer taxpayers left of notch-point and positive values represent more taxpayers. For nearly every year after the introduction of the credit there is a statistically significant positive estimate, averaging 4,032 more taxpayers just below the credit qualifying income cutoff. The only year for which there is no significance is 2009. While the exact reason for this year's insignificance is not certain, one clue is in the unusual uptick in the number of taxpayers who met all eligibility requirements and made a retirement contribution in that year relative to the surrounding years, presented earlier in Table 1.2. Around this same time the Bush Administration mailed stimulus checks through the Bush Economic Stimulus Package and the Obama Administration cut withholdings as part of the American Recovery and Reinvestment Act, all in the hopes of stimulating the economy during the Great Recession. The interactions of these two fiscal policies is the most likely underlying reason for 2009's insignificance, but further research would be required.

Figure 7.3 plots these estimates over time. The absence of bunching prior to 2002 serves to validate the assumption that bunching is the result of the Saver's Credit introduction. Of note, the estimated bunching effects are largest for 2002-2006 which perfectly correspond to both the years for which previous Saver's Credit studies investigated credit impact and the years for which the credit was not yet indexed to inflation. This would suggest that previous analyses may have overestimated the long run behavioral impact of income manipulation to qualify by way assuming the first five years' trends could correctly represent the trends going forward. But more importantly, it also suggests that inflation indexing may have the unintended secondary benefit of making it harder for taxpayers to manipulate their income to perfectly maximize the credit benefit. The IRS announces Saver's Credit income cutoffs approximately a year and a half before the filing season commences for any particular year. In other words, the income cutoffs are announced several months before the start of a calendar year. For example, the cutoffs for the 2013 tax year where announced in the fall of 2012 and tax

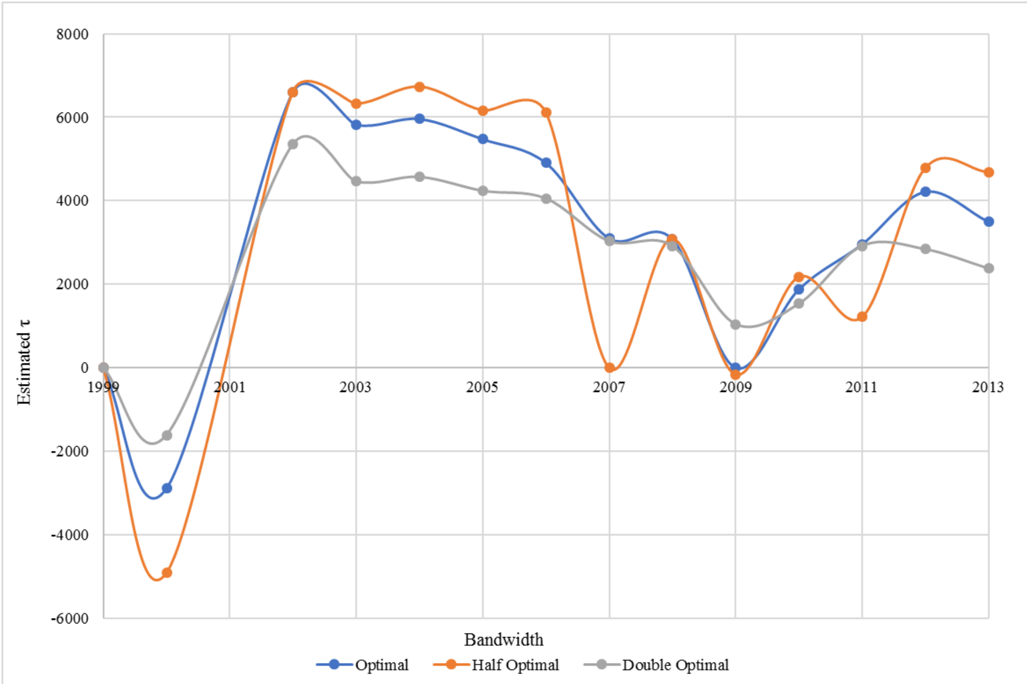


Figure 7.3: Estimated Number of Excess Taxpayers Below the Income Cutoff, 1999-2013

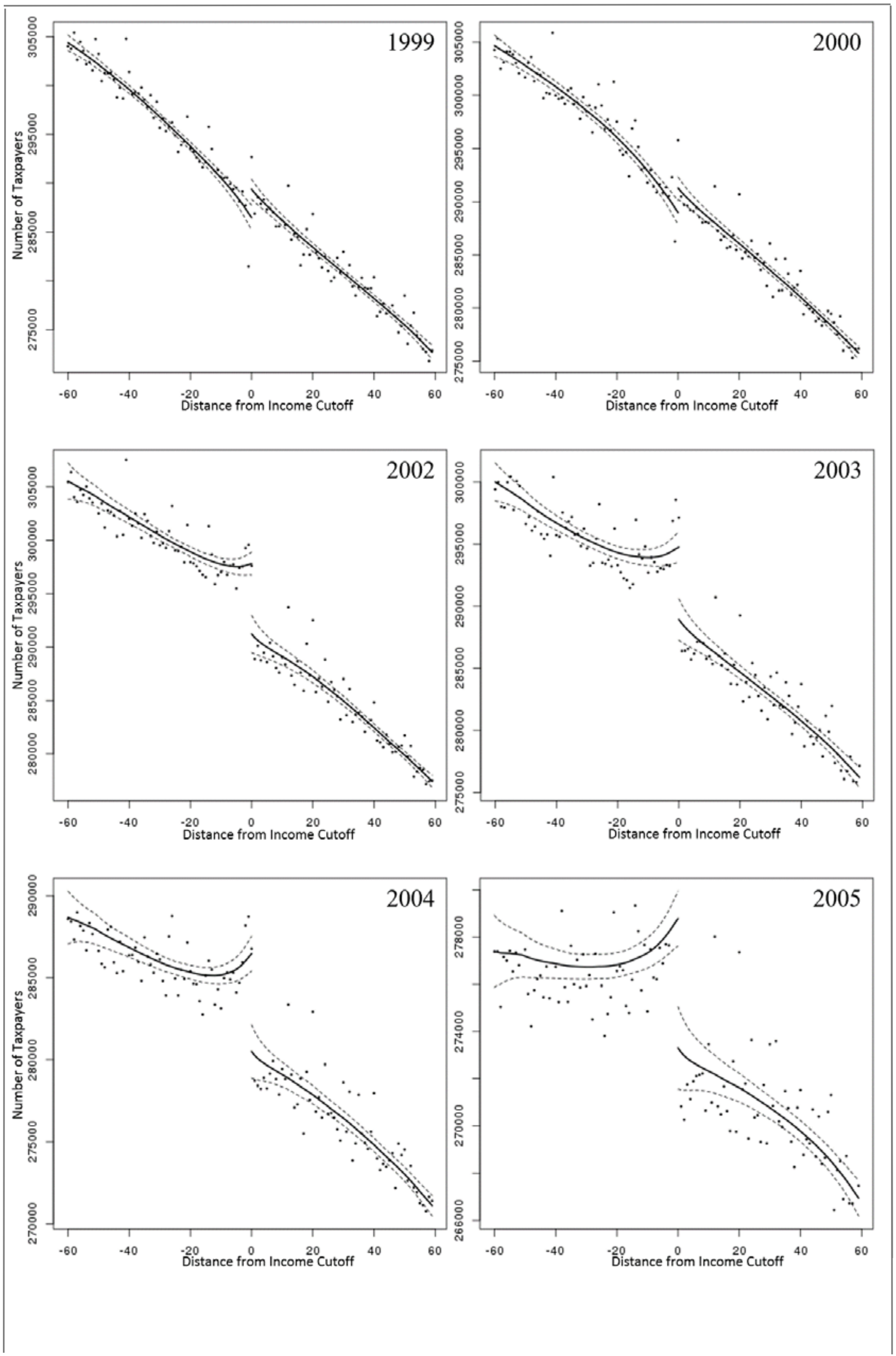
filings are due in the spring of 2014. This provided taxpayers the entire year to adjust income and savings behavior to maximize credit benefit. While true, this finding warrants further investigation.

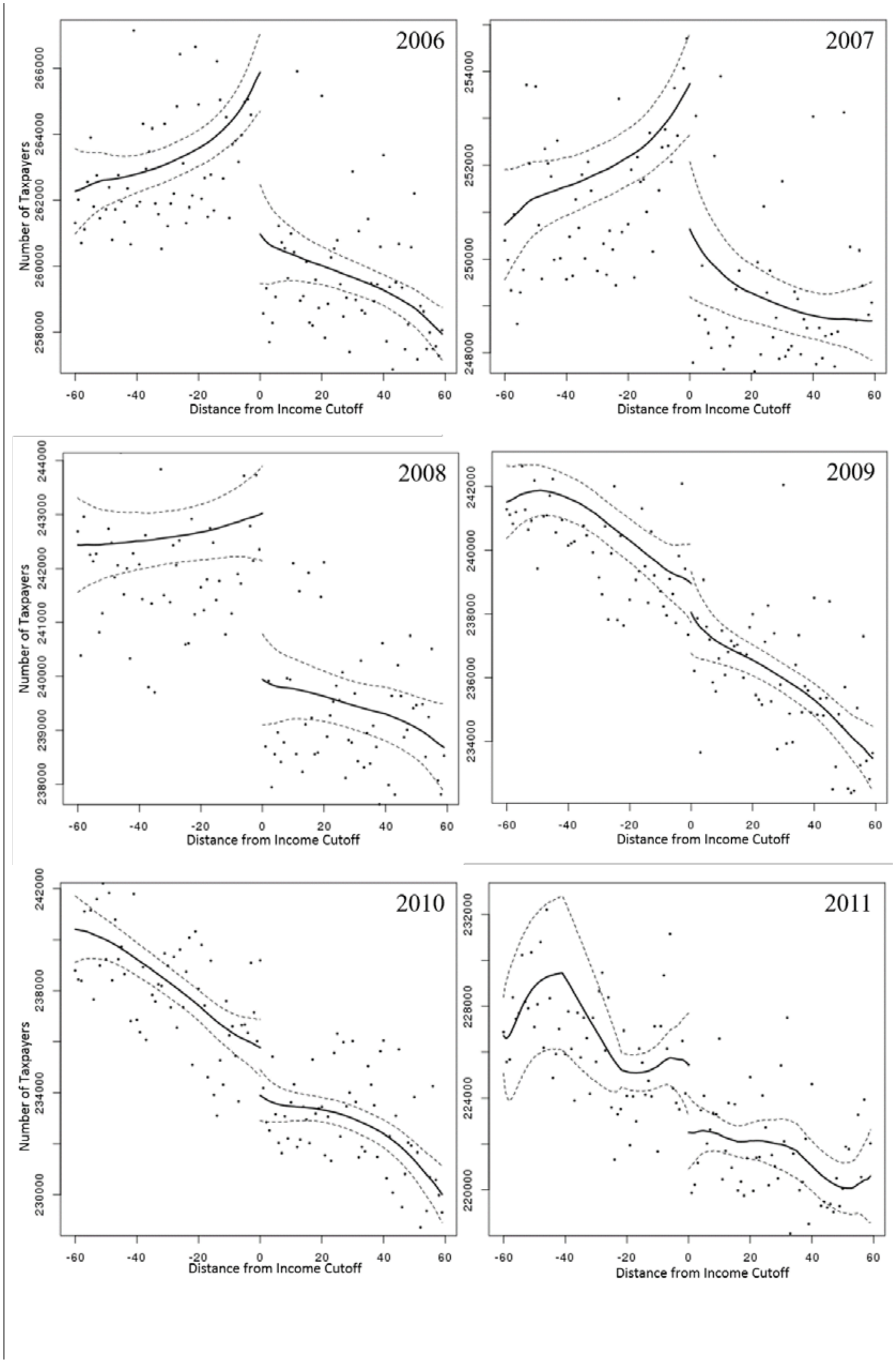
Another possible explanation could be the onset of the Great Recession that generally coincides with this decline. The National Bureau of Economic Research (NBER) estimated the recession officially began December of 2007 and lasted through June 2009 with full recovery by 2012. Looking at Figure 7.3, we see a corresponding dip in 2007 that remained low through 2012. However, bunching estimates remained below pre-recession rates in both 2012 and 2013. So while the recession may have contributed to the decline, it does not fully explain the patterns observed.

Figure 7.4 zooms in on each year's density distribution at the notch-point and overlays the estimated local linear regression. The distance between the left and right side regression lines equal the estimated τ reported in discontinuity measurement of Table 7.1. These graphics serve to highlight the discontinuity in density of reported AGI.

When broken out by the specific credit rate, described in Appendix G, we see that the 10 percent rate is the strongest driver of these effects. The 50 percent rate's estimated bunching impact is smaller and more consistent over time than that of the full dataset. The estimated bunching effect for the 20 percent is neither statistically significant nor consistently observed over time. The impact of the 10 percent rate mimics those of the full dataset and generally reports the largest estimated magnitudes. Recall, the Saver's Credit is structured to give the lowest rate to the highest income group. As highlighted in Chapter 6, there is a positive relationship between the size of AGI and making a retirement contribution, thus giving those near the 10 percent credit rate cutoff the largest incentive to adjust income in order to qualify.

This method of estimating the density discontinuity indicates a violation in the RD design's continuity in the running variable assumption. On average there is a discontinuity of 4,032 more taxpayers just left of the notch-point. While this confirms design assumption violations, it only estimates the difference in the number of taxpayers precisely at the notch-point and speaks nothing to the magnitude of bunching for regions further away from the notch-point. It is likely that taxpayers are incapable of perfectly adjust their AGI to fall exactly at the notch-point and end up locating themselves further away in the distribution from the notch-point than this method can account for. In these


1999
2000
2002
2003
2004
2005
Number of Taxpayers
Distance from Income Cutoff
-60
-40
-20
0
20
40
60
305000
295000
285000
275000
300000
290000
280000
270000
278000
274000
266000


2006
2007
2008
2009
2010
2011
Number of Taxpayers
Distance from Income Cutoff
-60
-40
-20
0
20
40
60
258000
260000
262000
264000
266000
248000
250000
252000
254000
238000
239000
240000
241000
242000
243000
244000
234000
236000
238000
240000
242000
230000
232000
234000
236000
238000
220000
224000
228000
232000

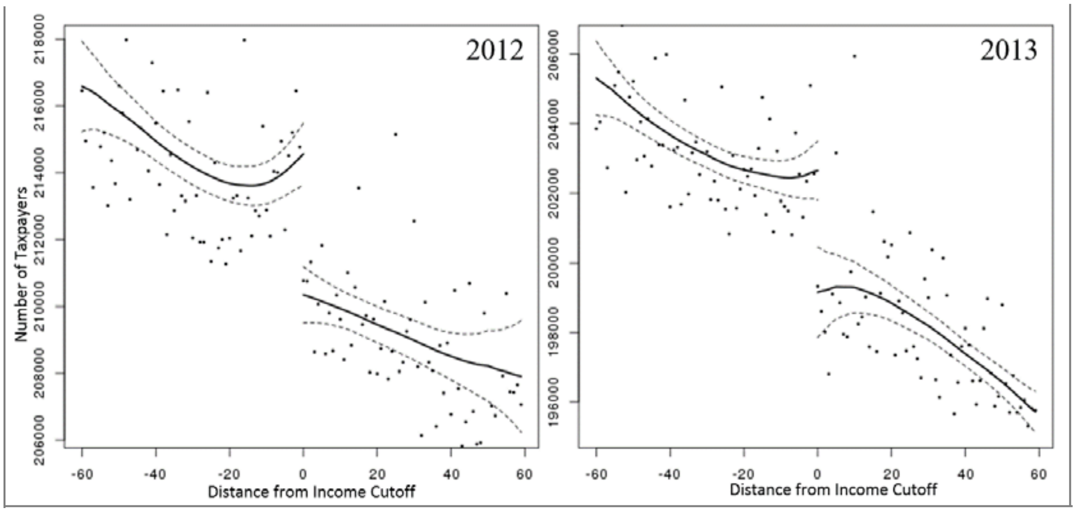


Figure 7.4: Density Discontinuity by Year

instances, this method of estimating discontinuity ignores such behavior, and consequently potentially underestimates the true magnitude of bunching. In the next section I expand the view to allow for this imprecise manipulation of income to better estimate the full magnitude of income manipulation.

Section 7.1.2 – Estimating bunching in density region

Because it is quite difficult to target income precisely it is inappropriate to assume the estimated discontinuity in the previous section fully captures the magnitude of bunching. To better estimate, I employ the standard approach developed by Chetty et al. (2011) and extended by Kleven and Waseem (2013). This requires first deriving the counterfactual density distribution in the absence of bunching by way of excluding taxpayers near the notch-point. I then use this to compare the observed density distribution to the computed counterfactual distribution to determine how many more taxpayers are located left of the notch-point than would be expected.

Specifically, I estimate the counterfactual density by fitting a polynomial *p*, to the distribution of taxpayers while excluding those in the bunching region *r*. The functional form is:

$$C_j = \sum_{i=0}^{p} \beta_i {Z_j}^i + \sum_{k=-r}^{r} \phi_k 1[Z_j \neq k] + \epsilon_j \tag{7.5}$$

where *j* represents bins of filers, C_j is the count of taxpayers in bin *j*, Z_j is the continuous distance from the notch-point but is set to zero when Z_j equals the excluded bunching region *r*, which was set to the optimal bandwidth derived from the underlying data. The best fit polynomial order was determined to be a 5th-order polynomial (*p*=5) for the underlying data.[26] This results in a counterfactual, $\widehat{C_j}$, equal to the predicted values from equation (7.5). The magnitude of bunching, then, is simply the difference between the observed cumulative density and the expected cumulative density. The difference between the observed C_j and counterfactual $\hat{C}_j$ below the notch-point represents the size of excess bunching. Figure 7.5 stylistically presents how this method works. Each blue and red dot represents actual taxpayers, binned to $50 increments away from the notch-point. Red dots indicate taxpayers falling within the *r* window and blue dots indicate taxpayers outside. The black line represents the polynomial regression derived from equation (7.5), which is omitted from the *r* window but is instead replaced by green dots. These green dots represent the counterfactual $\hat{C}_j$. Looking specifically to the left of the notch-point, the aggregated difference between the red and green dots determines the

[26] The underlying data was constrained to include only taxpayers with AGI $3,000 above and below the Saver's Credit income cutoff. This was done to ensure no overlap with other tax credit kink points or other Saver's Credit rate cutoffs that taxpayers may be responding to, which serve to bias both the model and subsequent results.

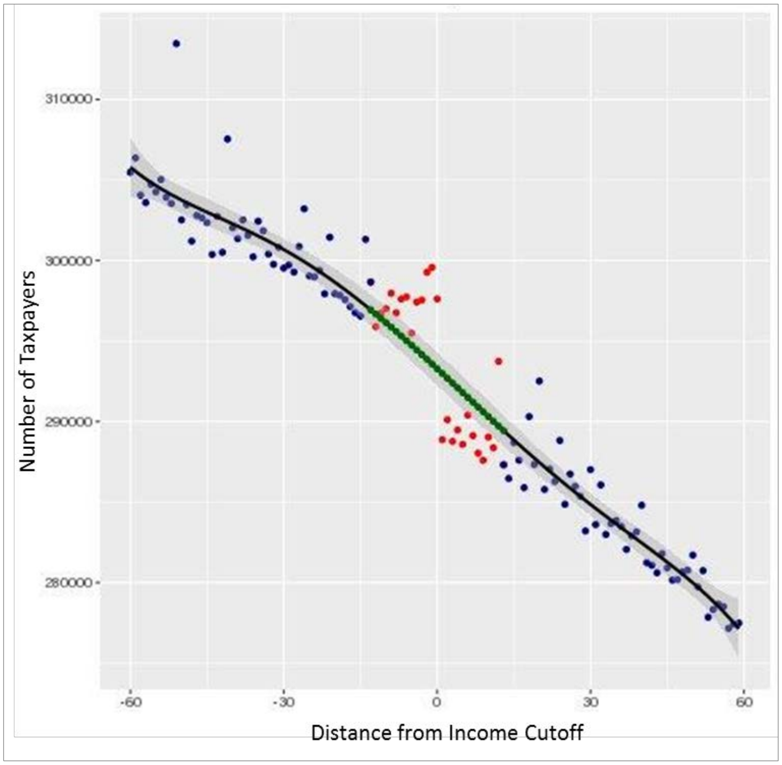


Figure 7.5: Stylized Distribution of Taxpayers with and without Counterfactual

excess bunching in the density window. The results presented here were robust to changes in *p* and *r*.

Table 7.2 reports the estimated size of bunching when compared to the estimated number of taxpayers in the absence of bunching. The first column represents the excess number of taxpayers. The second column is the corresponding percent of all taxpayers in the bunching region that were above (below) prediction. The first two years, 1999 and 2000, both indicate some degree of bunching, though the size is quite small and insignificant relative to years post credit availability. For years after the credit's introduction there is consistent and significant bunching of taxpayers above expectation,

Table 7.2: Estimated Number of Taxpayers Above Counterfactual, 1999-2013[1]

Year	Number of Taxpayers		Percent of all Taxpayers	
1999	7,838	***	0.087	***
2000	5,334	***	0.060	***
2002	33,227	***	0.419	***
2003	61,527	***	0.642	***
2004	36,114	***	0.474	***
2005	48,664	***	0.506	***
2006	40,341	***	0.375	***
2007	43,038	***	0.399	***
2008	34,040	***	0.266	***
2009	13,238	***	0.179	***
2010	-4,132	**	-0.045	**
2011	33,845	***	0.719	***
2012	52,673	***	0.637	***
2013	9,709	***	0.156	***

[1] Bandwidth is determined by employing Imbens-Kalyanamaran algorithm.
Signif. Codes: 0 '***' 0.001 '**', 0.01 '*', 0.05 '.' , 0.1 ' ', 1
Significance based on boot-strapped standard errors from 10,000 iterations.
Source: author's computation, Compliance Datawarehouse, IRS

except for 2010 and 2013[27] Because these estimates include regions further away from the exact notch-point it is unsurprising to find the size of bunching is significantly larger than previously estimated in Table 7.1. In fact, I find an average of 33,524 taxpayers bunching below the notch-point, compared to the 4,032 from the previous method. In general, these results confirm the findings from the previous subsection; bunching exists and is substantial for years after the introduction of the credit.

When comparing trends over time between the two methods, while the magnitudes are different, the patterns remain largely the same. The first several years

[27] For these years, see Appendix H.

show consistently large bunching proceeded by a precipitous decline and then general rebound, Figure 7.6. Here, the black line represents the estimated excess bunching estimate and the blue line represents the estimated density discontinuity, τ. Only three years appear to significantly deviate from each other, 2004, 2010, and 2013.[28]

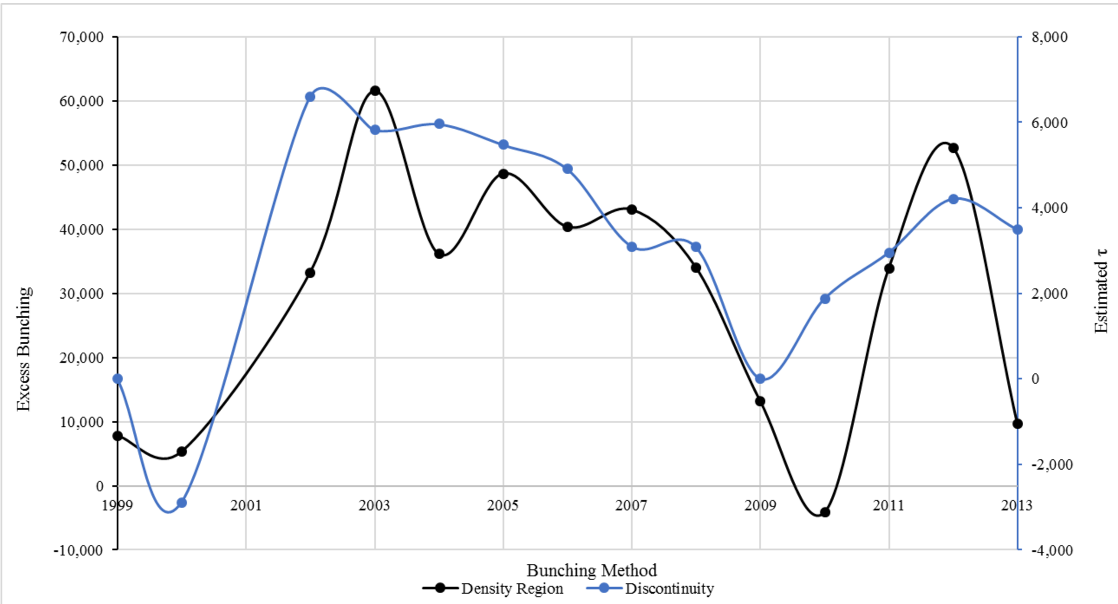


Figure 7.6: Comparative Estimated Number of Excess Taxpayers between Regression Density Discontinuity Method and Bunching Density Region Method, 1999-2013

By employing these two methods to examine the continuity of AGI across the notch-point, we can conclusively reject the RD assumption of a smooth running variable. The exact magnitude of bunching ranges from an average 4,032 taxpayers at the notch-point, to as much as an average 33,524 taxpayers if one expands the region of bunching

[28] It is unclear why 2004 deviates so substantially and calls for further research. The case for 2010 and 2013 are both made in Appendix H.

to areas past the notch-point. In either event, this presence of discontinuity calls into question the direction of causality. It is quite likely that taxpayers are choosing to manipulate their income to fall just below the notch-point in order to qualify for the Saver's Credit after having already decided to make a retirement contribution. This suggest a reverse in causality, with the Saver's Credit not inducing more retirement contributions but retirement contributions inducing credit activity. This ostensibly leaves us rejecting the RD design as a means to measuring the causal relationship. Such a However, recently several new approaches to combatting this model violation have been developed that can be employed to overcome this initial setback. I turn to those methods presently.

Section 7.2 – Regression Discontinuity with Adjustments – Bounding the Estimate

In the previous section I tested and rejected the RD assumption of a continuous density for filers around the notch-point. To overcome this setback the literature has developed numerous methods for dealing with this violation. I focus on one such method here.[29] This method, formally introduced in Sallee (2011) and used in Ramnath (2013), estimates a bounded treatment effect by removing extreme retirement contribution observations. I further develop this approach below.

To estimate the impact of the Saver's Credit while adjusting for discontinuities in the running variable, I estimate an upper and lower bound to the true treatment effect. To

[29] In Appendix I, I explore using 2001 income as a proxy for current year income. Another method available is the doughnut-hole approach in which one merely omits observations on either side of the notch-point from analysis (Hoxby and Bulman 2015). This latter approach calls into question the ability to assume those near the notch-point, now potentially much further away, are similar in all other regards. Given this limitation, I have chosen to ignore this approach in the research here.

understand how this works, one can think of taxpayers with income near the Saver's Credit notch-point as falling into one of two categories, income manipulators and non-manipulators. Income manipulators are taxpayers who have an incentive to adjust reported income in order to be just below the notch-point and therefore qualify to receive the credit. Income non-manipulators do not adjust their income and can fall either below or above the notch-point. Recall from equation (7.2), in the absence of income manipulation the treatment effect is the difference between retirement savings contributions below the notch, $E[A_i|D = 1]$, and contributions above, $E[A_i|D = 0]$. But, because taxpayers can manipulate income, $E[A_i|D = 1]$ is actually a combination of income manipulators, m, and non-manipulators, η, represented by

$$E[A_i|D = 1] - E[A_i|D = 0]$$

$$= \rho E[A_i|D = 1, m] + (1 - \rho)E[A_i|D = 1, \eta] - E[A_i|D = 0] \qquad (7.6)$$

where ρ is the percent of taxpayers who choose to manipulate their reported income and $(1 - \rho)$ is the percent of taxpayers who do not. Assuming income manipulators do not change their savings behavior then

$$\rho E[A_i|D = 1, m] = \rho(E[A_i|D = 1, \eta] - \varphi) \qquad (7.7)$$

where φ is the difference in mean retirement contributions between income manipulators and non-manipulators. I can therefore simplify (7.6) to

$$= [A_i|D = 1, \eta] - \rho\varphi - E[A_i|D = 0] \qquad (7.8)$$

where $\rho\varphi$ represents the bias introduced through income bunching. If φ is greater than zero, the treatment effect, τ from (7.2), becomes a lower bound estimate of the true effect[30], if φ is zero, τ represents the true effect of the Saver's Credit, and if φ is less than zero, τ is an upper bound estimate. We cannot estimate φ directly nor can we assume income manipulators always save more, or less, than their non-manipulating counterparts. However, we can make assumptions that will allow us to bound the true effect.

To accomplish this, one needs to produce an estimate that first necessarily undervalues τ and then necessarily overvalues τ. As discussed in Sallee (2011) and Horowitz and Manski (1995), two assumptions need to be applied.

Assumption 7.1: Equal population distribution for non-manipulators to the left and right of the notch point

$$n_{\eta}^{D=0} = n_{\eta}^{D=1} \tag{7.9}$$

If n_{η} represents the size of the population of non-manipulators to the right, D=0, and to the left, D=1, of the notch-point, then this assumption states that the size of the distribution of non-manipulators to the left and right of the notch-point are equal. What this suggests is that the swell of taxpayers to the left of the notch-point and above the expected non-manipulator density line identifies the number of all income manipulators. Because the full population to the left of the notch, $n^{D=1}$, is observable and the population to the right of the notch is both observable and populated by only non-manipulators $n_{\eta}^{D=0}$, then the number of income manipulators can be estimated as:

$$n_{m}^{D=1} = n^{D=1} - n_{\eta}^{D=0} \tag{7.10}$$

[30] This is due to the negative sign in front of $\rho\varphi$. This mathematically represents the dampening of the true $E[A_i|D = 1]$.

What this does not determine is the distribution of retirement contributions between manipulators and non-manipulators. To bound the effect, I need to isolate and remove income manipulators. This leads to a second assumption.

Assumption 7.2: Complete asymmetry in retirement contributions between manipulators and non-manipulators

For a lower bound estimate: $min[A_i|D = 1, m] \geq \max[A_i|D = 1, \eta]$

For an upper bound estimate: $min[A_i|D = 1, \eta] \geq \max[A_i|D = 1, m]$

This assumption states that to produce a lower bound estimate the minimum retirement contribution by income manipulators, $min\ [A_i|D = 1, m]$, is as large or larger than the maximum contribution by non-manipulators, $max\ [A_i|D = 1, \eta]$. The converse is assumed to produce an upper bound estimate. This is perhaps an unreasonable assumption in practice, but its purpose is to represent the most extreme potential difference between the two types of taxpayers. This allows for the identification of income manipulators as simply those contributing the largest/smallest amounts toward retirement savings in the observed distribution. Once the size, $n_m^{D=1}$, and contribution distribution is identified, I simply remove the income manipulators.

Mechanically, I sort all taxpayers by the size of their retirement contributions. For the upper bound, I identify all taxpayers starting at the taxpayer with the smallest contribution through the n^{th} taxpayer, where $n = n_\eta^{(D=0)}$. These taxpayers are then removed from the RD model represented in equation (7.1). Conversely for the lower bound, I flip the ordering so as to identify all taxpayers with the largest contributions through the n^{th} taxpayers and remove them from the RD model.

The combination of these two assumptions produces both a lower and an upper bound on the true treatment effect. Table 7.3 shows the results of this exercise. The lower bound estimates reflect the effect of removing taxpayers with the largest retirement contributions and the upper bound estimates reflect the effect of removing taxpayers with the smallest retirement contributions. The true estimate of the effect of the Saver's Credit lies somewhere in between. Form this, the Saver's Credit has had a marginally positive impact on average retirement contributions of between $55.12 (lower bound) and $83.64 (upper bound) per taxpayer per year.

Table 7.3: Bounded Effect of the Saver's Credit on Size of Retirement Contributions by Size of Credit Rate, 2002-2013

Year	Overall effect		Size of Credit					
			50% credit		20% credit		10% credit	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
2002	42.03 *	67.74 ***	47.02 *	85.24 ***	-8.72	28.81 ***	23.83	85.60 ***
2003	73.37 ***	87.53 ***	97.25 ***	102.59 ***	23.53 **	32.47 ***	49.21	114.17 ***
2004	70.88 **	87.77 ***	59.06 *	103.23 ***	16.73	32.88 ***	59.87	118.79 ***
2005	59.92 **	101.70 ***	89.78 ***	127.09 ***	27.47 **	38.93 ***	69.91 *	129.04 ***
2006	59.69 **	99.35 ***	110.27 ***	111.55 ***	23.69 *	40.92 ***	48.74	136.00 ***
2007	53.10 **	83.23 ***	70.50 ***	100.67 ***	-0.64	39.92 **	34.06	110.91 ***
2008	65.41 ***	80.50 ***	52.45 *	94.73 ***	-0.48	20.69 ***	49.26	114.01 ***
2009	41.70 *	67.90 ***	43.79 *	100.28 ***	1.78	23.89 **	62.77 ***	65.60 ***
2010	17.26	79.78 ***	33.63	92.88 **	-5.45	48.79 ***	38.54 *	85.82 ***
2011	72.35 ***	81.84 ***	103.22 ***	114.79 ***	-20.31	29.16 ***	51.83 *	91.73 ***
2012	60.32 ***	79.51 ***	68.07 **	119.52 ***	15.05	15.60 ***	60.99 **	84.03 ***
2013	45.37 .	86.78 ***	74.71 **	143.70 ***	8.51	13.56 .	86.00 ***	91.68 ***

[1] Bandwidth determined by employing Imbens-Kalyanamaran algorithm.
Signif. Codes: 0 '***' 0.001 '**', 0.01 '*', 0.05 '.' , 0.1 ' ', 1
Source: author's computation; Compliance Datawarehouse, IRS

Interestingly, barring 2002, the first several years show the credit having the largest effects. Starting in 2007, the impact of the credit begins to decline precipitously, coinciding with the start of the Great Recession and the first year of AGI indexing, the rules of which were discussed in more detail in Chapter 2. There appears a resurgence in

2011, with an effect ranging between $72.35 and $81.84, but never reaches the level of influence it accomplished in earlier years. Figure 7.7 highlights the range of possible true credit effects on contributions spanning the twelve-year study window. In every year except 2010, the true effect is bounded between values greater than zero. From this, we can conclude that the Saver's Credit has had a positive impact on retirement contributions, albeit small.

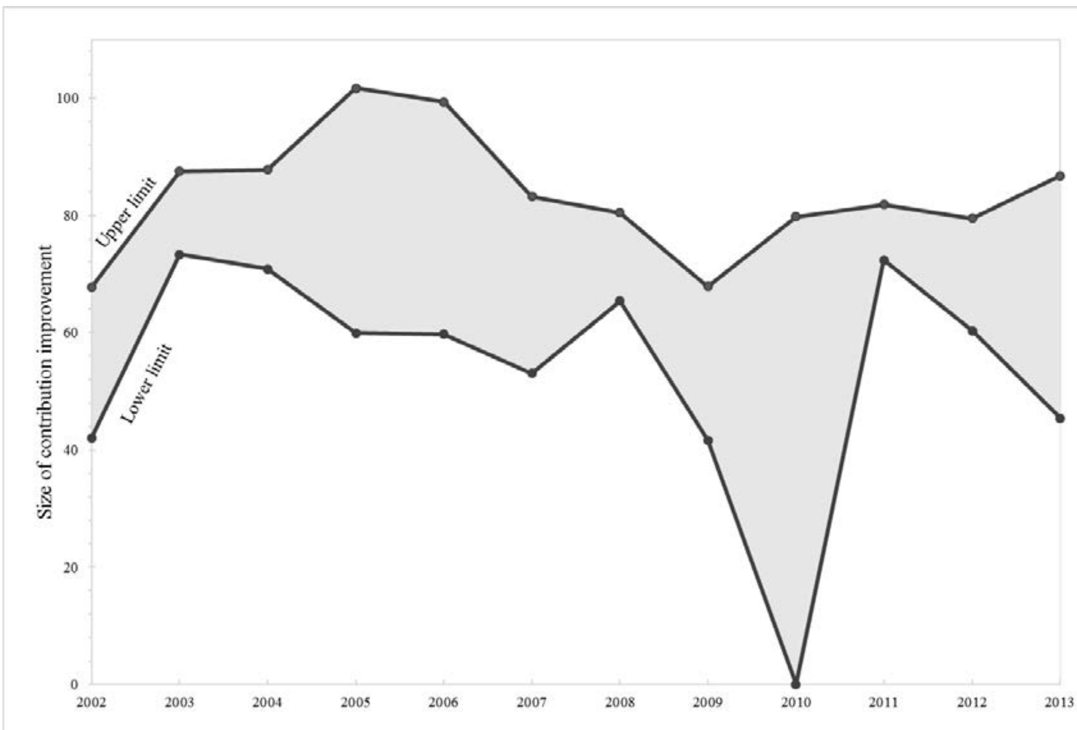


Figure 7.7: Bounded Effect of the Saver's Credit on Retirement Contributions, 2002-2013

Table 7.3 also breaks these bounded estimates down by credit rate. To better visualize these ranges, Figure 7.8 replicates Figure 7.7 but subdivides by credit rate. For the 50 percent rate, the Saver's Credit marginally improved contributions by an average of between $70.81 and $108.02. These effects are on the high end of the various credit

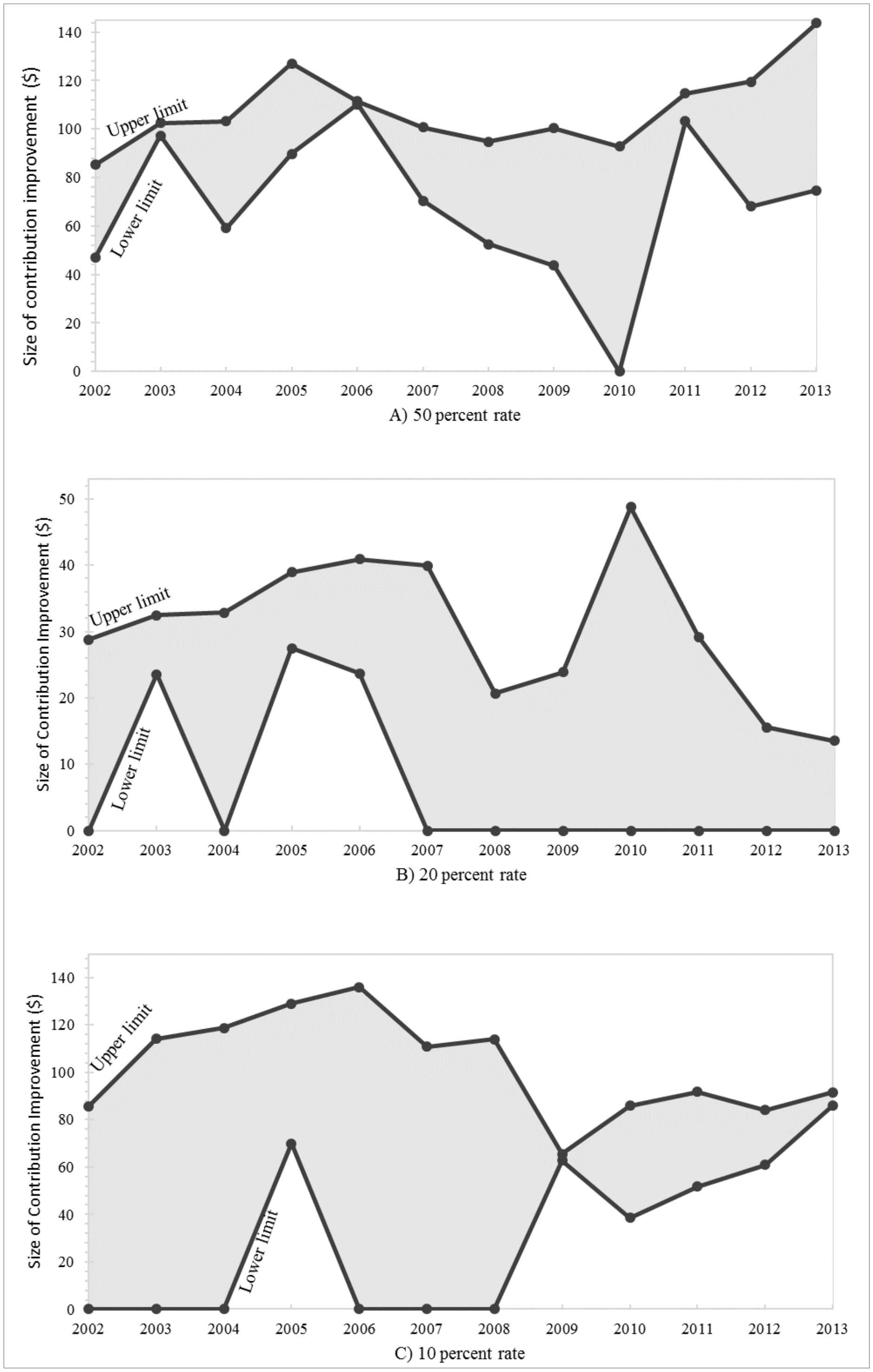


Figure 7.8: Bounded Effect of the Saver's Credit on Retirement Contributions by Size of Credit Rate, 2002-2013

rate responses, suggesting that the 50 percent rate had the most influence on incentivizing new retirement contributions. The 20 percent rate, on the other hand, improved contributions an estimated average of between $6.76 and $30.47 and were rarely proven to be statistically significantly different than zero. This implies that the 20 percent rate had little to no effect on taxpayer contribution behavior.

The 10 percent rate, conceptually, presents the largest incentive to adjust behavior. This is due to the higher likelihood of taxpayers' capable of making retirement contributions given higher income levels and, more importantly, the binary determination of qualifying for any Saver's Credit benefits. While 10 percent is the least generous credit rate offered, this binary determination is not present for either of the higher credit rates as having higher or lower income affords you some Saver's Credit benefits. As an example, having income just above the 20 percent credit rate income threshold still qualifies one for the 10 percent credit rate benefit. However, having income just above the 10 percent credit rate income threshold disqualifies the taxpayer from receiving any credit benefits. Though the case may be, the impact of the 10 percent rate is on average lower than the 50 percent rate, with an estimated average improvement of contributions between $52.92 and $102.28. Unlike the other rates, the 10 percent rate is not initially statistically significant for the lower bound but becomes significant in later years. This implies that the 10 percent rate has become an increasingly important factor in taxpayers' decisions surrounding retirement contributions. This is further supported by the narrowing range of the possible true treatment effect.

From these estimates it is clear the Saver's Credit has had an impact on improving retirement contributions. Though the exact range and size of that impact depends on the rate and year examined, overall the credit has proven to be successful in encouraging new contributions.

Section 7.3 – Discussion

In this study, unlike the previous three studies, I go beyond merely exploring participation to examine whether the Saver's Credit induced a change in taxpayer behavior. Specifically, if the Saver's Credit impacted retirement contributions as is its intention. I first tested and proved violations of smoothness in the distribution of income across the notch-points, for all years after the credit's introduction. Then, in dealing with this RD design assumption violation, I intentionally biased the model to produce a bounded treatment effect, which encompasses the true effect.

The detection of bunching in the income distribution is alone an important finding. This definitively concludes that the Saver's Credit has affected taxpayer behavior, albeit not in the intended way. The persistence of bunching, though declining, which starts in the first year of credit availability, is evidence of a taxpayer response to the Saver's Credit. While the intent was to induce retirement contributions, the mere existence of bunching is evidence that the credit is creating behavioral change.

There is yet another important discovery that arises from this bunching exercise, the size of bunching precipitously declines just after the credit rate thresholds are indexed, in 2007. While certainly the Great Recession, the Economic Stimulus Act of 2008, and the American Recovery and Reinvestment Act of 2009, all played some part in this decline, the lack of a full recovery following these various events suggests that indexing the cutoff thresholds also played a role. Indexing, while implemented in order to account for inflation, may have had the unintended benefit of making income manipulation that much harder to attain. For policy makers this finding is quite illuminating. The purpose of a tax credit is generally to incentive preferred financial or social behavior, the fact that taxpayers are manipulating income to circumvent these incentives is naturally undesirable. The discovery that the simple act of indexing a credit's cut-points adversely affects a taxpayer's ability to skirt the intended response is very useful. Though this investigation was focused on measuring the impact of the Saver's Credit, this finding warrants further research.

The existence of bunching ostensibly nullifies the ability to utilize the RD design. In the presence of income bunching, the direction of impact potentially reverses. However, I overcome this setback by bounding the true treatment effect. If one assumes that income manipulators are the largest contributors to retirement plans, then omitting these taxpayers from the analysis results in a lower bound, as the taxpayers that remain represent the lowest contribution levels. If one assumes that income manipulators are the lowest contributors, then omitting these taxpayers from the analysis produces an upper bound, as now taxpayers that remain are the largest contributors. The true treatment effect then lies somewhere in between. By employing this approach, I find that the Saver's Credit has led to marginal increases in retirement contributions, ranging from an average of approximately \$55 to \$84. Overtime, this range remains relatively consistent, barring 2010. Subdividing by credit rate, the 50 percent rate has proven to have the largest impact on retirement contributions, followed closely behind by the 10 percent rate. The 20 percent rate appears to have no economic impact on retirement contributions, suggesting that its existence serves to afford little more benefit than complicating the incentive structure.

Comparing these results against previous studies, my estimates fall on the higher side but with a much smaller variance, that does not include zero. Duflo et al. and Ramnath both estimate the impact to be somewhere between improving contributions by \$81 or \$127, respectively, or lowering contributions by as much as \$32 or \$356, respectively. My research, with its enhanced data quality, concludes that the Saver's Credit has resulted in improvements in contributions. Perhaps somewhat surprisingly, while participation has grown over time, the bounded impact of the credit on contributions has not followed suit. Going forward, it would be beneficial to further subdivide the population by those characteristics that afford more (less) opportunities for income manipulation. In so doing, we could better to understand the interplay between manipulation capabilities and credit impacts.

Overall, this research shows that the Saver's Credit has had marginal improvements on retirement contributions, focused almost entirely on the 10 and 50 percent credit rate. These marginal improvements are on the order of 12 to 15 percent increase, for the lower bound, and 19 to 24 percent increase, for the upper bound, in retirement contributions from the previous average contribution rate of approximately \$470. For future research, it would be beneficial to further explore which types of contributions, deferred compensation or IRA contributions, are the most affected.

While these results clearly indicate both a change in taxpayer behavior due to the Saver's Credit, and improvements in retirement contributions, there are several future directions and unanswered questions that remain. I have already highlighted a few such as subdividing the population based on filing characteristics and isolating the type of retirement contribution made. In addition, it would be important to explore the robustness of these results to smaller and larger bin sizes. The analysis here assumed that taxpayers anchored income reporting to $50 increments. However, it may well be the case that taxpayers cluster around $100 or even $1,000 increments. If this is so, then these estimates may not be fully capturing the density, or contributional change, around the notch-points.

CHAPTER 8 – CONCLUSIONS

This research has shed light on the previously unexplored gradual structural change in Saver's Credit participation across the U.S. and, more specifically, the Southeast region along with incorporating how that gradual change has influenced taxpayer response. It examines where and how participation spreads and quantifies the magnitude and marginal change over time. It also examines the direction of influence, whether the credit has driven retirement contributions or if contributions have driven credit participation. This project marries the four fields of spatial dependence, federal tax expenditure interactions with local jurisdictions, Saver's Credit participation, and causal analysis for the first time. It both exposes and fills in the gaps of our understanding of how taxpayers react and respond to federal tax law incentives.

I utilized three different spatial approaches for examining the flow and movement of Saver's Credit participation across numerous populations of interest. In the first, I took a visual thematic approach to explore the movement, density, and overlap of participation with retirement contribution behavior. Such an approach highlights the varied density distributions underlying national statistics on the take-up of the Saver's Credit. I showed the Appalachian region and areas in and throughout Kentucky present with the highest rates of participation regardless of the population of interest explored. I also showed that making retirement contributions are heavily concentrated in urban areas and throughout the Appalachian. Additionally, I exposed the limitation of shorter time horizons that prior studies of the Saver's Credit have employed. I presented evidence that by the end of the

15-year study window, nearly all areas have expanded participation and retirement contributions. I also showed this expansion occurring rather consistently year-over-year radiating out from the initially dense areas.

In the second investigation, I expanded on the apparent radiating effect observed in the first study to explicitly measure this spatial dependence. Through the application of the Moran's I, I exposed the consistent but diminishing spatial correlation in credit participation and the constant correlation in retirement contribution participation. Such findings confirm the existence of a spatial component impacting taxpayer behavior which previous research generally overlooked. Beyond simply exposing correlation, I went further to specifically identify highly clustered areas of high/low participation. Confirming the visual pattern of the first study, high participation areas are strongly spatially correlated throughout the Appalachian, throughout Kentucky and Tennessee, while low participation areas are strongly spatially correlated among urban areas and throughout Florida.

Having established the existence of a spatial relationship in the earlier two studies, my third investigation takes these findings even further to directly measure the factors driving participation along with explicitly measuring the size and direction of spatial spillovers contained therein. I first showed the error in applying a traditional OLS model when spatial dependence is expected, in the form of either inefficient standard errors or biased coefficient estimates. I then derived three different regression models that explicitly address spatial relationships. Through their application I proved the unsuitability of the OLS model in favor of the SDM. I consistently found that spillovers drive most, if not all, of the observed total effect with the three largest parameters presenting with spillover effects being Married filing jointly, making a retirement contribution, and using a paid preparer. I also showed that these spillover effects diminish over time, which is consistent with results from the earlier two studies.

Finally, I turned to addressing the impact of the Saver's Credit on taxpayer behavior. I first highlighted the presence and magnitude of reverse causality through two varying methods. First, I measured the density discontinuity at the income notch-points, this resulted in an estimated average 4,032 more taxpayers just below the credit qualifying income cutoff. Given that taxpayers are often unable to perfectly adjust income to fall at the notch-point, I then produced an estimated bunching size by computing the difference between the observed density distribution and a predicted counterfactual distribution across the notch-point. I found that an average of 33,524 taxpayers bunched below the income threshold to qualify for Saver's Credit benefits. I then adjust the traditional RD design to control for this bunching by producing a range of Saver's Credit treatment effects. To do this, I derived an upper and lower bound adjustment by making two assumptions about the underlying income manipulation behavior of the filing population. In so doing, I estimate that the Saver's Credit has had an overall effect of increasing retirement contributions by as little as an average $55 or as much as an average $84. When broken down by each credit rate, the 50 percent rate had the largest impact on contributions by a ranging averaging between $70.81 and $108.02, the 10 percent rate had the next largest impact averaging between $52.92 and $102.28, and the 20 percent rate had very little impact ranging between $6.76 and $30.47.

Overall, the findings from this research are significant to our understanding of how taxpayers respond to federal tax policy. I revealed a very strong spatial dynamic underpinning the observed take-up of the Saver's Credit. We now definitively know that credit participation is heavily dependent on location, as are the drivers influencing participation. We also now know that the growth in credit awareness and utilization radiates out from initially highly clustered areas and that the impact of spatial dependence diminishes as awareness improves. Moreover, we now know that the Saver's Credit has, at least marginally, had the intended impact of increasing annual retirement contributions. The results from this research can be directly applied to informing policy makers on the patterns around tax expenditure activity growth and how that growth organically expands. This research serves to identify where targeted outreach would be most effective. The radiating effect observed in Chapter 4 and corroborated and quantified in Chapters 5 and 6 serve to highlight the most advantageous localities for targeted outreach. Specifically, targeting outreach to paid preparers in areas with consistently highly correlated low participation would serve to be the most cost-efficient means for improving participation.

Given the large absence in the literature exploring spatial spillovers' effects on federal tax filing behavior, this research also serves as a roadmap for future investigations into the impact that spillovers have on tax policy. This methodology can be employed on a number of other federal, more salient, tax expenditures to compare and contrast the spatial influences underpinning the observed patterns of filing behavior. For example, such expenditures could include the Earned Income Tax Credit or the Child Tax Credit.

While illuminating, this investigation leaves some questions unanswered and introduces more. I focused much of my attention on the Southeast region, but it is unclear if the patterns observed for this area hold into other regions of the U.S. I focus entirely on measuring spatial dynamics on credit participation, but it is unclear whether these same conclusions can be drawn for credit amount claimed. Additionally, subdividing the population based on filing characteristics and isolating the type of retirement contribution made would further our understanding on how to minimize the adverse response to the Saver's Credit and maximize improvements to retirement contributions.

APPENDIX A – THE GEARY'S C AND GETIS-ORD G STATISTICS

While the Moran's I statistic is the workhorse for estimating and establishing spatial autocorrelation, there are still two other tests that can address spatial relationship; the Geary's C and the Getis-Ord G tests. While both can prove to be useful, the former suffers from a slightly less than intuitive interpretation, relative to the Moran's I, and the latter suffers from a lack of precision that is provided by the Moran's I. The Geary's C test is, in application, the nearly exact inverse of the Moran's I. Where Moran's I measures the similarity of values across neighboring areas, Geary's C measures the difference across those areas. It is defined as

$$C = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{ij}\left(z_i - z_j\right)^2}{2S_z^2\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}}$$

The values range between 0 and 2, with a mean expected value of 1. Values less than 1 are positively correlation and values greater than 1 are negatively correlated. Given its similarity in computation to the Moran's I, we should expect to see a pattern of positive similarity that declines over time.

The Getis-Ord G statistic does not measure spatial autocorrelation directly but rather spatial association. In other words, this measures the clustering of values. It is defined as

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{ij}z_i z_j}{\sum_{i=1}^{n}\sum_{j=1}^{n}z_i z_j}$$

G analyzes across the whole of the distribution. Thus, distributions with many high and low clusters will have a canceling effect on the statistic, potentially resulting in the incorrect interpretation of no spatial association. Spatial autocorrelation measures avoid this error. When the spatial weight matrix is binary, as is the case here, the resulting *G* will range between 0 and 1. Interpretation however, depends on the value of the z-score, a positive z-score implies high values are clustered and a negative z-score implies low values are clustered, relative to expected randomness. From the previous study, observing the patterns in high participation areas makes it easy to expect areas of high value clustering. There also appears strong clustering of low values as well. Thus, we expect *G* to suffer from this canceling out effect to result in outcomes of little to no clustering patterns.

Table A.1 lists out the results of these two additional statistics. As expected, *C* follows the same pattern as *I*. *C* is consistently below 1, indicating more spatial similarity than would be expected, and the impact declines over time. This holds for both participation in the Saver's Credit and retirement contributions. The results of *G* are also generally as expected. It is somewhat surprising to find retirement contribution participation is statistically significant for all eligibility groups but only statistically significant for Subgroup 1 and a couple of years of Subgroup 2 with respect to Saver's Credit participation. The z-score for retirement contribution participation indicates a clustering of positive values but both as time passes and we more precisely target our population of interest, its strength declines. For the only statistically significant eligibility group with respect to Saver's Credit participation, the z-score also indicates a clustering of positive values and also declines over time.

Table A.1: Spatial Association Statistics by Eligibility Group, Geographic Region, and Activity of Interest, 2002-2013

	Geary's C					Getis-Ord G [1]									
	Saver's Credit Participation			Retirement Contribution Participation		Saver's Credit Participation						Retirement Contribution Participation			
						Population		Subgroup 1		Subgroup 2		Population		Subgroup 1	
Year	Population	Subgroup 1	Subgroup 2	Population	Subgroup 1	G estimate	z-score	G estimate	z-score	G estimate	z-score	G estimate	z-score	G estimate	z-score
Continental U.S.															
2002	0.5787 ***	0.5851 ***	0.7431 ***	0.6589 ***	0.6580 ***	0.0247		0.0253		0.0248		0.0281 ***	50.73	0.0258 ***	8.47
2003	0.5874 ***	0.5972 ***	0.7630 ***	0.6458 ***	0.6454 ***	0.0245		0.0252		0.0248		0.0281 ***	49.45	0.0258 ***	7.01
2004	0.5932 ***	0.6141 ***	0.7994 ***	0.6397 ***	0.5667 ***	0.0241		0.0251		0.0248		0.0280 ***	49.81	0.0257 ***	5.86
2005	0.6070 ***	0.6158 ***	0.8070 ***	0.6394 ***	0.6551 ***	0.0244		0.0254		0.0251		0.0279 ***	48.45	0.0258 ***	6.22
2006	0.6161 ***	0.6333 ***	0.8293 ***	0.6304 ***	0.6500 ***	0.0244		0.0255 *	1.78	0.0252		0.0280 ***	50.44	0.0258 ***	7.46
2007	0.6244 ***	0.6337 ***	0.8179 ***	0.6204 ***	0.6463 ***	0.0244		0.0255 *	1.54	0.0253		0.0277 ***	45.33	0.0257 ***	5.78
2008	0.6570 ***	0.6503 ***	0.8583 ***	0.6206 ***	0.6475 ***	0.0247		0.0255 *	2.20	0.0254		0.0275 ***	43.53	0.0256 ***	4.44
2009	0.6618 ***	0.6548 ***	0.8549 ***	0.6169 ***	0.6348 ***	0.0249		0.0256 ***	3.49	0.0254 *	2.27	0.0275 ***	42.81	0.0256 ***	5.03
2010	0.6594 ***	0.6532 ***	0.8614 ***	0.6202 ***	0.6459 ***	0.0247		0.0255 *	2.50	0.0255 ***	4.60	0.0275 ***	41.68	0.0255 ***	3.27
2011	0.6757 ***	0.6740 ***	0.8689 ***	0.6176 ***	0.6499 ***	0.0247		0.0255 *	2.31	0.0255 ***	4.03	0.0274 ***	39.46	0.0256 **	3.02
2012	0.6947 ***	0.6851 ***	0.8600 ***	0.6147 ***	0.6611 ***	0.0247		0.0256 **	2.60	0.0255 ***	3.30	0.0273 ***	37.74	0.0256 ***	3.96
2013	0.6974 ***	0.6893 ***	0.8752 ***	0.6129 ***	0.6577 ***	0.0248		0.0256 **	3.00	0.0255 **	3.03	0.0272 ***	36.84	0.0257 ***	4.34
Southeast Region															
2002	0.6004 ***	0.7214 ***	0.9270 ***	0.6162 ***	0.7775 ***	0.0110		0.0113		0.0117		0.0134 ***	28.37	0.0117	
2003	0.6215 ***	0.7330 ***	0.9566 *	0.6069 ***	0.7519 ***	0.0110		0.0114		0.0118		0.0134 ***	28.63	0.0117	
2004	0.6345 ***	0.7379 ***	0.9662 *	0.6078 ***	0.7731 ***	0.0109		0.0114		0.0117		0.0134 ***	29.17	0.0179	
2005	0.6356 ***	0.7591 ***	0.9341 ***	0.6318 ***	0.8043 ***	0.0110		0.0114		0.0117		0.0133 ***	28.56	0.0118	
2006	0.6230 ***	0.7635 ***	0.9375	0.6205 ***	0.7983 ***	0.0110		0.0115		0.0119		0.0133 ***	27.89	0.0118	
2007	0.6528 ***	0.8020 ***	1.0091	0.6065 ***	0.8339 ***	0.0108		0.0157		0.0119		0.0133 ***	27.68	0.0118	
2008	0.6264 ***	0.8060 ***	1.0153	0.6739 ***	0.8191 ***	0.0109		0.0115		0.0119		0.0131 ***	25.62	0.0173	
2009	0.6408 ***	0.7919 ***	1.0137	0.6827 ***	0.7905 ***	0.0110		0.0115		0.0119		0.0130 ***	25.11	0.0117	
2010	0.6442 ***	0.7837 ***	0.9815	0.6997 ***	0.8010 ***	0.0111		0.1165		0.0120		0.0130 ***	24.24	0.0118	
2011	0.6819 ***	0.8401 ***	1.0494	0.6964 ***	0.8119 ***	0.0110		0.0116		0.0120		0.0130 ***	22.91	0.0118	
2012	0.7258 ***	0.8858 ***	1.0694	0.7199 ***	0.8155 ***	0.0110		0.0116		0.0120		0.0129 ***	23.49	0.0118	
2013	0.7349 ***	0.8855 ***	1.0934	0.7062 ***	0.8167 ***	0.0110		0.0117		0.0119		0.0129 ***	23.61	0.0119	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

[1] Z-score shown only for statistically significant G estimates

Source: author's computation; Compliance Datawarehouse, IRS

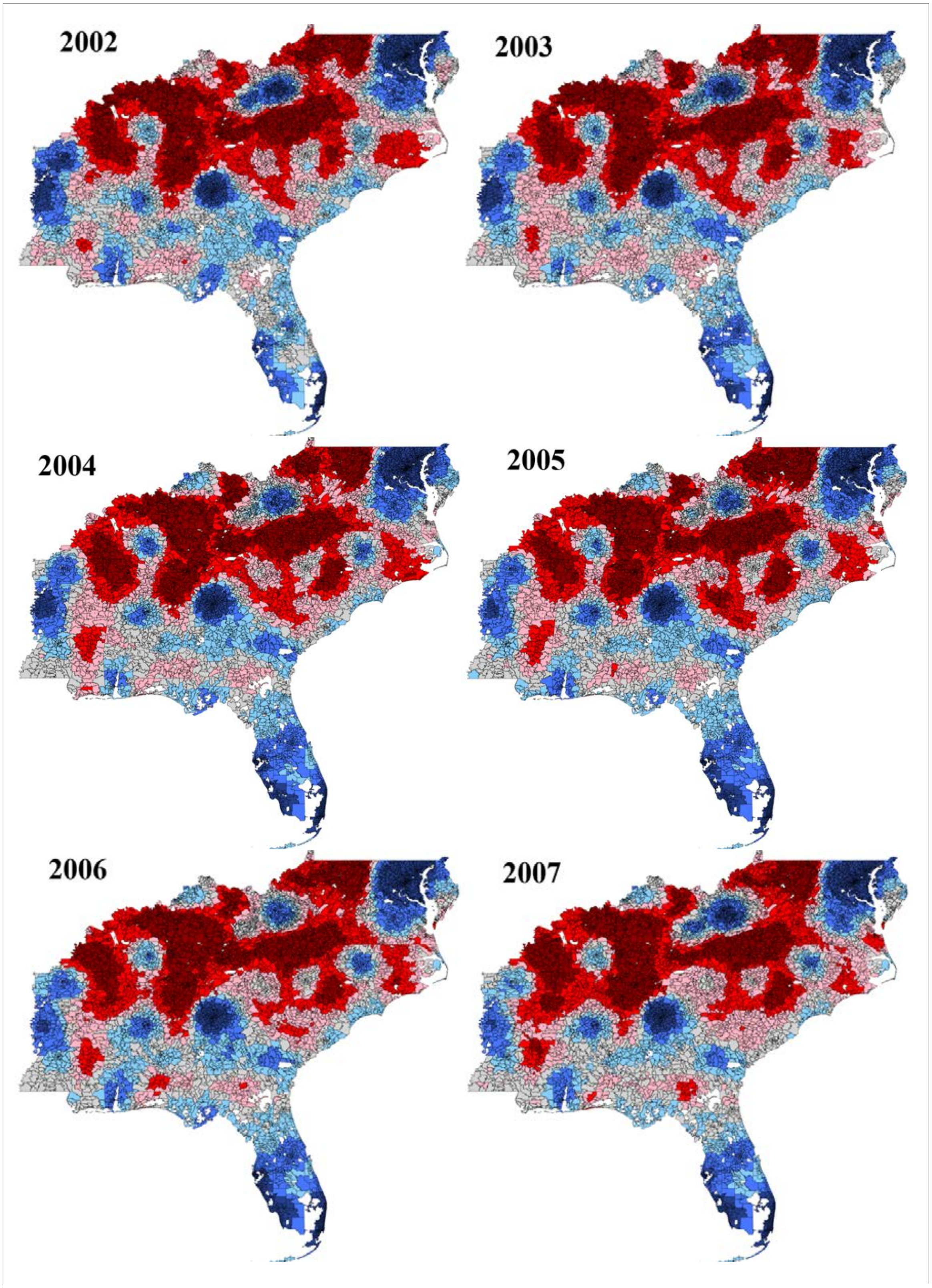
APPENDIX B – THE GETIS-ORD G_i^* STATISTIC

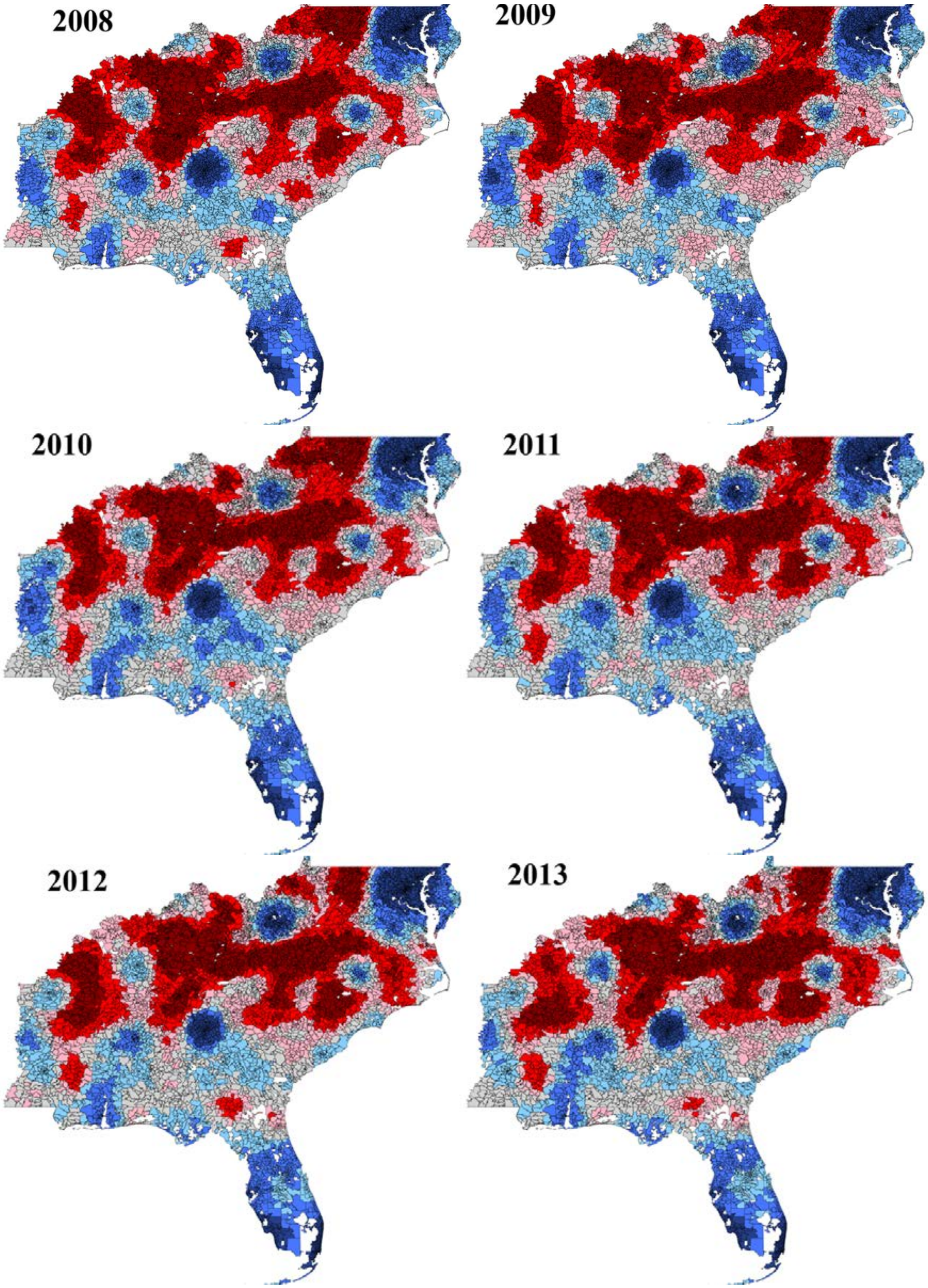
While not definitionally a LISA, the Getis-Ord G_i^* analyzes the spatial association of like valued areas. Where the local Moran's I identified statistically significant locations of high correlation, G_i^* offers the ability to locate spatial clusters of high values and spatial clusters of low values independently. Computationally, it is defined as

$$G_i^* = \frac{\sum_{j=1}^{n} \omega_{ij} x_i - \bar{X} \sum_{j=1}^{n} \omega_{ij}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} \omega_{ij}^2 - \left(\sum_{j=1}^{n} \omega_{ij} \right)^2 \right]}{n-1}}}$$

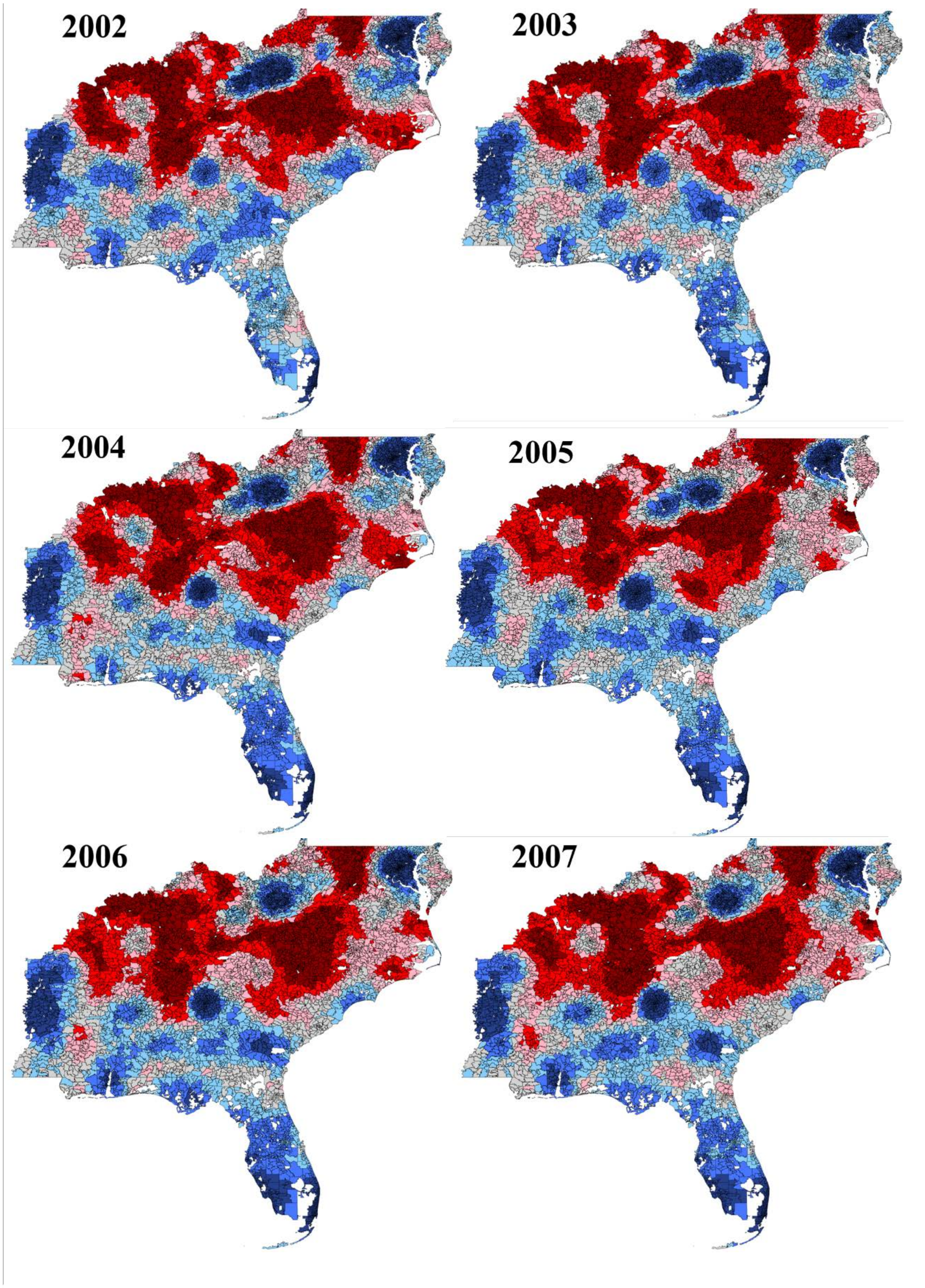
Without going into the exact derivation, the G_i^* is essentially the decomposition of the Getis-Ord *G* for location *i* while concurrently deriving the z-score. The benefit of this statistic is the ability to not only pinpoint locations of spatial association but to determine the strength of that association. The resulting map is typically coined a Hot Spot analysis because of its ability to identify hot and cold spots of clustered activity. This statistic is a compliment to the local Moran's I.

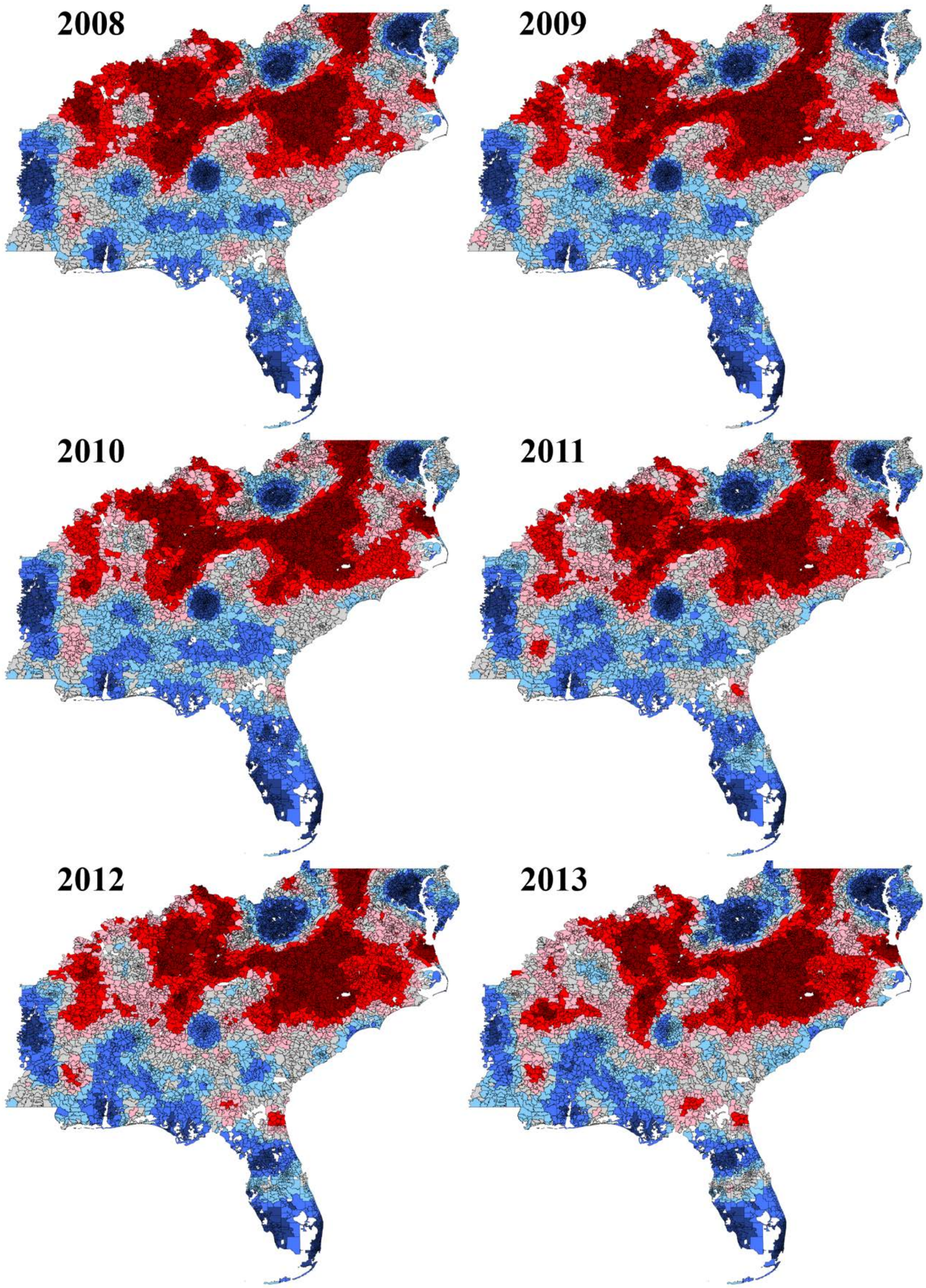
Figures B.1.A-C present the Hot Spot maps of Saver's Credit participation. Those areas in shades of red identify local hot spots for high participation while areas shaded in blues identify local cold spots, or clusters of low participation. Similar to results in Figures 5.3.A-C, spatial clusterings of high activity occur in and throughout the Appalachian areas, through Kentucky, Tennessee, and on North Carolina's coast. Spatial clustering of low activity is concentrated in urban and surrounding areas. These maps offer a slightly more robust view than their local Moran's I counterpart in that one can now observe the peaks and nadirs of the distribution more easily.


2002
2003
2004
2005
2006
2007

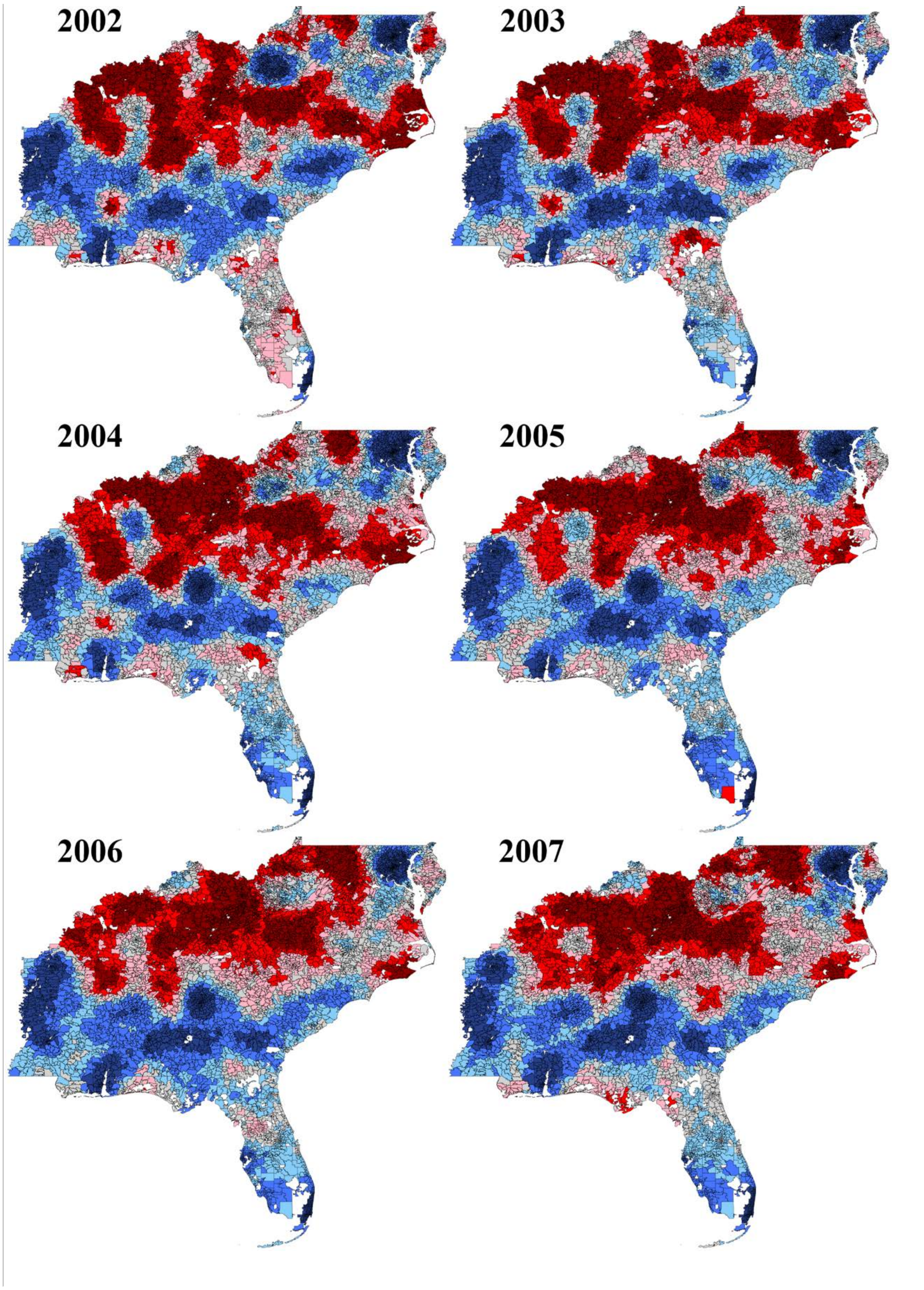


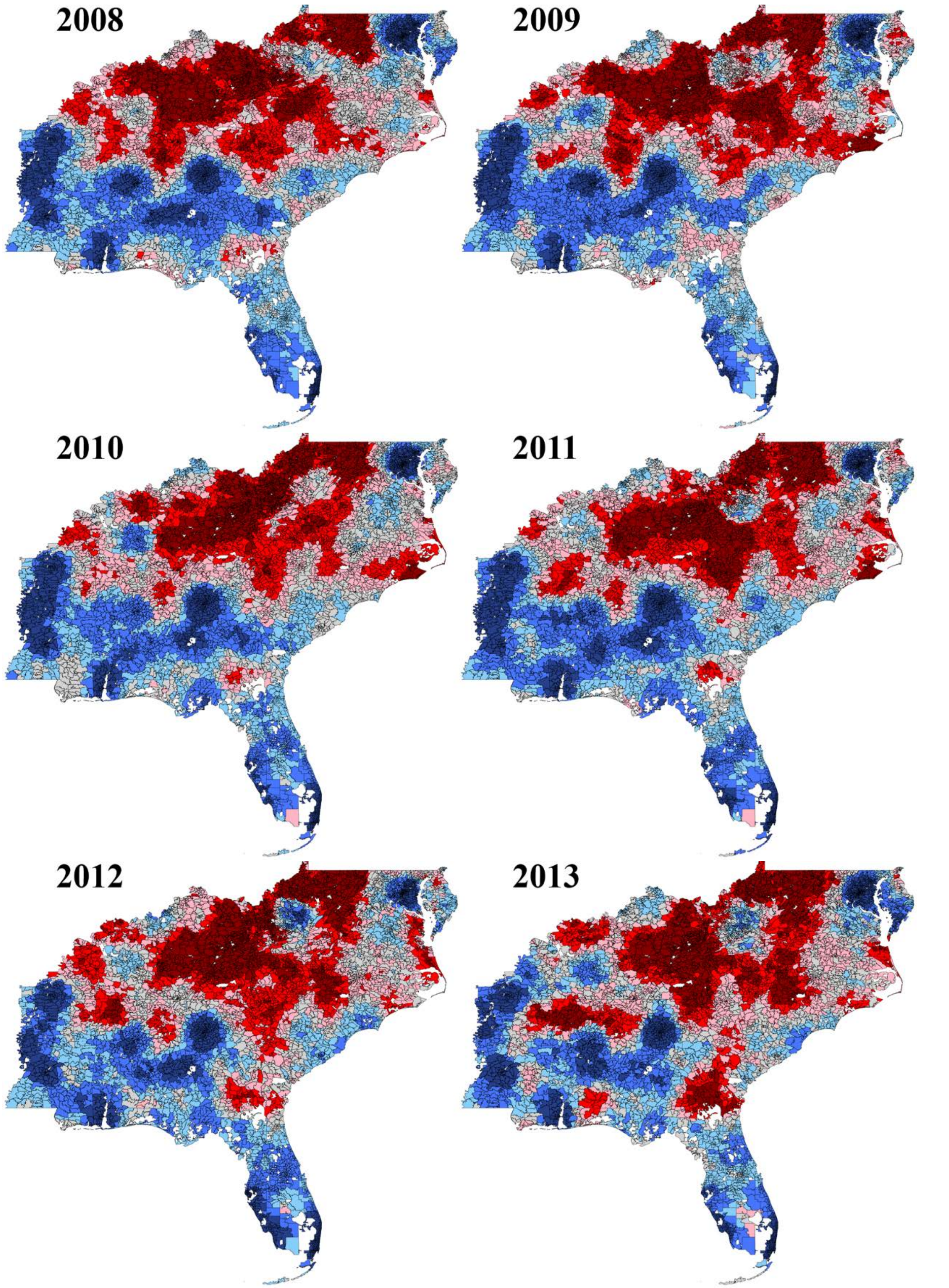
A) Full Population


2002
2003
2004
2005
2006
2007



B) Subgroup 1


2002
2003
2004
2005
2006
2007



C) Subgroup 2

Figure B.1: Hotspot Map of Saver's Credit Participation by Eligibility Group, 2002-2013

APPENDIX C – STUDY POPULATION COMPARED TO CENSUS POPULATION

For this research I focused exclusively on information that was obtainable from IRS administrative data systems. This results in the exclusion of pertinent Census information for which detailed demographics could not be attained. There are two overarching reasons I chose to omit Census information. First, demographic data at the 5-digit ZIP Code level was not available for every year of my study. Second, the populations covered by my study were not comparable to population demographics provided by Census. I justify this decision to exclude Census data here.

This study is based on the 5-digit ZIP Code reported on the 1040. Census produces demographic information at that same level of detail through either decennial publications and estimates from the American Community Survey (ACS). Matching Census data with IRS data at the 5-digit ZIP Code level could provide for added demographic controls such as industry, race, education, ethnicity, and unemployment. However, Census only offers this level of granularity for 2000 and 2010 forward. This leaves demographic data entirely unobserved for years 2002-2009.

An alternative would be to geographically overlay Census' County data and ascribe county percentages to the various ZIP Codes that are included in each county. This offers the benefit of providing some demographic information for years there otherwise would not be. However, this too is limited. First, Census only provides individual demographic data at the county level for years 2005 and later. This still leaves 2002-2004 unaccounted for.

In addition to not covering all years, imposing county level estimates to the underlying ZIP Codes is not ideal from either the modeling perspective or the spatial coverage perspective. It is unclear whether distributions observed at the county level are evenly distributed across included ZIP Codes. It may be the case that a county reports an *X* percent of manufacturing jobs but the distribution of those jobs is entirely concentrated in one or a few ZIP Codes to which this percent is ascribed. This would result in estimates that were perhaps even more misleading than having excluded it entirely. To further explain; if one credited, say, *X* percent of manufacturing jobs for all underlying ZIP Codes, the Lagrange Multiplier test would erroneously estimate a highly correlated spatial relationship of those jobs across ZIP Codes and thus reject the OLS model in favor of a spatial model, when this correlation was entirely of our own creation. A similar concern would arise in the coefficient estimates derived from the spatial models. And still more, any resulting coefficient based on county level data would be attenuated and thus, biased due to the significant measurement error embedded within. From a practical standpoint, it is unclear how one would handle ZIP Codes that straddle county lines. Does the ZIP Code in question take an average percentage estimate obtained from each of the counties it straddles, or some other tie breaking method?

Assuming for a moment that Census provided ZIP Code level data for every year of this study, it is still not clear that their inclusion is entirely merited. Recall, the population of interest in this study is limited to taxpayers between the ages of 18 and 57 in 1999 and reporting AGI below $100,000. Census makes none of these same stipulations. One needs only look at the distribution of population counts between Census

Table C.1: Population Count Comparison between Study Population and Census Population, 2010
(Counts in Thousands)

Selected Demographics	Number of Taxpayers	Census Population Counts[1]
Population Size	110,410	245,645
Single	34,887	46,673
Married	58,053	121,840
Age < 35	26,358	71,163
Age 35 < 45	28,490	41,716
Age 45 < 55	28,943	44,784
Age 55 and older	26,618	75,151
Male	53,079	119,542

[1] Census counts are for age 15 years and older.

Source: Compliance Datawarehouse, IRS and CPS estimates, Census

and this study for 2010 to see a distinct difference in population size estimates, Table C.1. The population covered by this study is only 45 percent of the population Census estimates for that same year. With respect to marital status, this study covers 48 percent of all married persons and 75 percent of singles. It would perhaps be sufficient if the distributions among these demographics were in line, however, here too, there are major discrepancies. This study reports 48 percent of taxpayers are male while Census finds 49 percent. This study shows 24 percent of the population are 35 or younger, while Census reports 29 percent. It is clear these two populations are not the same and so to attribute Census derived demographic estimates to this study muddies the extracted impacts observed in the model.

One may be comfortable overlooking the mismatch in population, but as eligibility is further targeted, such allowances become increasingly harder to maintain.

Subgroup 1 represents only 14 percent of Census' population and Subgroup 2 only 3 percent. It is unreasonable to assume the Census ZIP Code demographic distributions represent these targeted populations when these groups only represent a mere fraction of the population. Any estimates derived from Census information would most certainly suffer from selection bias, the exact size of which would be unknown.

Given the lack of accurate information covering all years and the misalignment in population definition, I have chosen to not include Census demographic data in this analysis.

APPENDIX D – FULL SPATIAL MODEL RESULTS

Tables D.1-D.3 report model estimates disaggregated by year.

Tables D.4-D.6 report spatial direct, indirect, and total effects for the SDM model by year.

Table D.1: Full Population Model Estimates, 2002-2013 [1]

Parameter [2][3]	2002 OLS	2002 SAR	2002 SER	2002 SDM	2003 OLS	2003 SAR	2003 SER	2003 SDM	2004 OLS	2004 SAR	2004 SER	2004 SDM
Intercept	0.0000	0.0099	0.1154	0.0093	0.0000	0.0090	0.1017	0.0073	0.0000	0.0091	0.0965	0.0082
	(0.0080)	(0.0072)	(0.1060)	(0.0070)	(0.0077)	(0.0069)	(0.1128)	(0.0066)	(0.0077)	(0.0069)	(0.1001)	(0.0067)
Making Retirement Contributions	1.0110	0.7487	0.9006	0.9069	0.8658	0.6155	0.7847	0.7920	0.8138	0.5865	0.7465	0.7542
	(0.0489) ***	(0.0443) ***	(0.0464) ***	(0.0464) ***	(0.0461) ***	(0.0420) ***	(0.0440) ***	(0.0441) ***	(0.0451) ***	(0.0411) ***	(0.0427) ***	(0.0428) ***
Squared(Making Retirement Contributions)	-0.7331	-0.4870	-0.5345	-0.5353	-0.5476	-0.3157	-0.3571	-0.3576	-0.5240	-0.3079	-0.3539	-0.3547
	(0.0491) ***	(0.0442) ***	(0.0458) ***	(0.0458) ***	(0.0461) ***	(0.0417) ***	(0.0429) ***	(0.0430) ***	(0.0453) ***	(0.0410) ***	(0.0421) ***	(0.0421) ***
Total Adjusted Gross Income	-0.1365	-0.0998	-0.1122	-0.1095	-0.2076	-0.1753	-0.1809	-0.1788	-0.1995	-0.1644	-0.1656	-0.1646
	(0.0125) ***	(0.0113) ***	(0.0117) ***	(0.0117) ***	(0.0220) ***	(0.0198) ***	(0.0204) ***	(0.0204) ***	(0.0223) ***	(0.0200) ***	(0.0207) ***	(0.0207) ***
Squared(Total Adjusted Gross Income)	0.0587	0.0399	0.0424	0.0393	0.0928	0.0964	0.0978	0.0974	0.0831	0.0863	0.0860	0.0860
	(0.0104) ***	(0.0093) **	(0.0091) **	(0.0092) **	(0.0198) ***	(0.0178) **	(0.0177) ***	(0.0177) **	(0.0200) ***	(0.0180) **	(0.0179) **	(0.0179) **
Qualifying for 20 Percent Credit	0.1001	0.0750	0.0755	0.0769	0.1457	0.1100	0.1101	0.1121	0.1510	0.1131	0.1125	0.1128
	(0.0093) ***	(0.0084) ***	(0.0082) ***	(0.0082) ***	(0.0088) ***	(0.0080) ***	(0.0077) ***	(0.0078) ***	(0.0090) ***	(0.0081) ***	(0.0079) ***	(0.0079) ***
Qualifying for 10 Percent Credit	0.4317	0.3369	0.3648	0.3606	0.4581	0.3771	0.4046	0.4017	0.4501	0.3744	0.3953	0.3937
	(0.0106) ***	(0.0098) ***	(0.0100) ***	(0.0100) ***	(0.0099) ***	(0.0092) ***	(0.0093) ***	(0.0093) ***	(0.0098) ***	(0.0091) ***	(0.0093) ***	(0.0094) ***
90 percent or more of income from SE [4]	0.0419	0.0378	0.0374	0.0353	0.0450	0.0468	0.0470	0.0466	0.0358	0.0424	0.0396	0.0384
	(0.0094) ***	(0.0084) **	(0.0100) **	(0.0100) **	(0.0088) ***	(0.0079) ***	(0.0093) **	(0.0093) **	(0.0088) ***	(0.0079) **	(0.0094) **	(0.0094) ***
Male	-0.0297	-0.0062	-0.0007	0.0016	-0.0009	0.0200	0.0287	0.0305	0.0281	0.0428	0.0530	0.0540
	(0.0100) **	(0.0090)	(0.0089)	(0.0089)	(0.0097)	(0.0087) .	(0.0086) **	(0.0086) **	(0.0100) **	(0.0090) **	(0.0089) ***	(0.0089) ***
Between 25 and 34	-0.0730	-0.0364	-0.0718	-0.0706	-0.0636	-0.0343	-0.0824	-0.0810	-0.0382	-0.0297	-0.0776	-0.0755
	(0.0117) ***	(0.0105) **	(0.0107) ***	(0.0107) ***	(0.0122) ***	(0.0110) *	(0.0111) ***	(0.0111) ***	(0.0141) **	(0.0127) .	(0.0129) ***	(0.0129) ***
Between 35 and 44	-0.1618	-0.0823	-0.1052	-0.1038	-0.1450	-0.0846	-0.0957	-0.0940	-0.1065	-0.0674	-0.0817	-0.0800
	(0.0109) ***	(0.0099) ***	(0.0100) ***	(0.0100) ***	(0.0102) ***	(0.0093) ***	(0.0093) ***	(0.0093) ***	(0.0103) ***	(0.0093) ***	(0.0093) ***	(0.0094) ***
Between 45 and 54	-0.1123	-0.0770	-0.1263	-0.1244	-0.1008	-0.0712	-0.1206	-0.1189	-0.0602	-0.0528	-0.1014	-0.0995
	(0.0130) ***	(0.0117) ***	(0.0123) ***	(0.0123) ***	(0.0121) ***	(0.0109) ***	(0.0112) ***	(0.0112) ***	(0.0125) ***	(0.0113) **	(0.0115) ***	(0.0115) ***
Between 55 and 59	-0.0979	-0.0721	-0.0718	-0.0722	-0.0814	-0.0661	-0.0671	-0.0675	-0.0922	-0.0908	-0.0972	-0.0974
	(0.0126) ***	(0.0113) ***	(0.0113) ***	(0.0113) ***	(0.0128) ***	(0.0115) ***	(0.0113) ***	(0.0113) ***	(0.0131) ***	(0.0117) ***	(0.0116) ***	(0.0116) ***
Married filing Jointly	0.5820	0.3401	0.3040	0.2998	0.5238	0.3115	0.2433	0.2394	0.5445	0.3322	0.2993	0.2928
	(0.0297) ***	(0.0270) ***	(0.0287) ***	(0.0287) ***	(0.0291) ***	(0.0265) ***	(0.0282) ***	(0.0282) ***	(0.0293) ***	(0.0268) ***	(0.0284) ***	(0.0284) ***
Head of Household	0.3177	0.2262	0.2835	0.2870	0.2532	0.1855	0.2088	0.2101	0.3118	0.2359	0.2865	0.2866
	(0.0369) ***	(0.0331) ***	(0.0358) ***	(0.0358) ***	(0.0360) ***	(0.0323) ***	(0.0348) ***	(0.0348) ***	(0.0360) ***	(0.0324) ***	(0.0350) ***	(0.0351) ***
1 to 3 dependents	-0.1353	-0.0707	-0.0661	-0.0644	-0.1088	-0.0688	-0.0802	-0.0793	-0.0922	-0.0627	-0.0763	-0.0757
	(0.0174) ***	(0.0156) **	(0.0164) **	(0.0164) **	(0.0169) ***	(0.0152) **	(0.0159) **	(0.0159) **	(0.0170) ***	(0.0153) **	(0.0160) **	(0.0160) **
4 or more dependents	-0.0332	0.0123	0.0008	0.0012	0.0357	0.0539	0.0487	0.0491	0.0004	0.0161	0.0051	0.0061
	(0.0289)	(0.0258)	(0.0275)	(0.0275)	(0.0282)	(0.0253) .	(0.0267)	(0.0267)	(0.0278)	(0.0250)	(0.0265)	(0.0266)
Using a paid preparer	-0.2401	-0.0530	-0.1385	-0.1555	-0.4768	-0.1110	-0.2107	-0.2214	-0.6011	-0.1371	-0.2497	-0.2573
	(0.0584) ***	(0.0525)	(0.0630) .	(0.0635) .	(0.0581) ***	(0.0532) .	(0.0651) *	(0.0656) **	(0.0584) ***	(0.0541) .	(0.0664) **	(0.0671) **
Squared(Using a paid preparer)	0.2977	0.0886	0.2925	0.3087	0.5399	0.1588	0.3973	0.4071	0.6597	0.1870	0.4375	0.4460
	(0.0577) ***	(0.0520)	(0.0605) **	(0.0611) **	(0.0577) ***	(0.0530) *	(0.0623) ***	(0.0629) ***	(0.0581) ***	(0.0540) **	(0.0635) ***	(0.0643) ***
ρ		0.6269		0.8121		0.5965		0.8002		0.5965		0.8015
		(0.0128) ***		(0.0213) ***		(0.0126) ***		(0.0221) ***		(0.0128) ***		(0.0223) ***
λ			0.9342				0.9413				0.9334	
			(0.0111) ***				(0.0103) ***				(0.0112) ***	

Footnotes at the end of table.

Table D.1: Full Population Model Estimates, 2002-2013, cont.

	2002				2003				2004			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Lag(Making Retirement Contributions)				-(0.6516)				-(0.6589)				-(0.8470)
				(0.1778) **				(0.1682) **				(0.1664) **
Lag(Squared(Making Retirement Contributions))				(0.4369)				(0.3982)				(0.5886)
				(0.1791) .				(0.1691) .				(0.1723) **
Lag(Total Adjusted Gross Income)				(0.1413)				(0.1901)				(0.1244)
				(0.0538) *				(0.0881) .				(0.0893)
Lag(Squared(Total Adjusted Gross Income))				-(0.1707)				-(0.1263)				-(0.0909)
				(0.0612) *				(0.0955)				(0.0954)
Lag(Qualifying for 20 Percent Credit)				(0.0956)				(0.1226)				(0.0301)
				(0.0590)				(0.0551) .				(0.0522)
Lag(Qualifying for 10 Percent Credit)				-(0.2950)				-(0.3346)				-(0.2134)
				(0.0453) ***				(0.0453) ***				(0.0460) **
Lag(90 Percent or more of income from SE[4])				(0.0484)				(0.0348)				(0.0435)
				(0.0255)				(0.0240)				(0.0233)
Lag(Male)				(0.0090)				-(0.0119)				-(0.0795)
				(0.0497)				(0.0496)				(0.0519)
Lag(Between 25 and 34)				(0.0323)				(0.1506)				(0.1844)
				(0.0571)				(0.0593) .				(0.0684) *
Lag(Between 35 and 44)				-(0.0002)				(0.0645)				(0.0681)
				(0.0488)				(0.0455)				(0.0451)
Lag(Between 45 and 54)				(0.1153)				(0.1661)				(0.1861)
				(0.0539) .				(0.0520) *				(0.0557) **
Lag(Between 55 and 59)				(0.0057)				(0.1065)				(0.1193)
				(0.0593)				(0.0613)				(0.0647)
Lag(Married filing Jointly)				-(0.0463)				-(0.1205)				-(0.2318)
				(0.0962)				(0.0918)				(0.0906) .
Lag(Head of Household)				-(0.2040)				-(0.2604)				-(0.3228)
				(0.1132)				(0.1051) .				(0.1053) *
Lag(1 to 3 dependents)				-(0.0031)				(0.1171)				(0.1028)
				(0.0596)				(0.0584) .				(0.0566)
Lag(4 or more dependents)				(0.0492)				(0.1156)				(0.1071)
				(0.0927)				(0.0874)				(0.0850)
Lag(Using a paid preparer)				(0.4052)				(0.4516)				(0.4386)
				(0.1435) *				(0.1477) *				(0.1508) *
Lag(Squared(Using a paid preparer))				-(0.5472)				-(0.6256)				-(0.6181)
				(0.1429) **				(0.1466) **				(0.1500) **

Footnotes at the end of table.

Table D.1: Full Population Model Estimates, 2002-2013, cont.

	2002				2003				2004			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
N	7,469	7,469	7,469	7,469	7,468	7,468	7,468	7,468	7,462	7,462	7,462	7,462
AIC	15,776	14,158	13,886	13,818	15,040	13,488	13,090	13,016	15,074	13,549	13,185	13,123
Moran I Residuals	0.1710 ***	0.0381 ***	-0.0015	0.0011	0.0381 ***	0.0449 ***	0.0000	0.0028	0.1687 ***	0.0449 ***	-0.0051	-0.0022
Log Likelihood		-7,057.97	-6,922.03	-6,870.17		-6,723.03	-6,523.87	-6,469.09		-6,753.71	-6,571.52	-6,522.46
Lagrange Multiplier												
LMerr	7,257 ***				6,945 ***				7,049 ***			
LMlag	3,104 ***				2,802 ***				2,799 ***			
RLMerr	4,861 ***				4,827 ***				4,906 ***			
RLMlag	708 ***				684 ***				656 ***			

Footnotes at the end of table.

Table D.1: Full Population Model Estimates, 2002-2013, cont.

	2005				2006				2007			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Intercept	0.0000	0.0088	0.0899	0.0079	0.0000	0.0084	0.0860	0.0077	0.0000	0.0087	0.0940	0.0085
	(0.0075)	(0.0068)	(0.0929)	(0.0066)	(0.0074)	(0.0068)	(0.0876)	(0.0066)	(0.0076)	(0.0070)	(0.0931)	(0.0067)
Making Retirement Contributions	0.6615	0.4932	0.6189	0.6261	0.6807	0.5281	0.6548	0.6629	0.7080	0.5137	0.7124	0.7221
	(0.0420) ***	(0.0386) ***	(0.0398) ***	(0.0399) ***	(0.0424) ***	(0.0391) ***	(0.0405) ***	(0.0406) ***	(0.0441) ***	(0.0409) ***	(0.0419) ***	(0.0419) ***
Squared(Making Retirement Contributions)	-0.3388	-0.1841	-0.1980	-0.1987	-0.3398	-0.2105	-0.2255	-0.2279	-0.3396	-0.1880	-0.2584	-0.2615
	(0.0421) ***	(0.0385) **	(0.0390) **	(0.0391) **	(0.0425) ***	(0.0389) **	(0.0399) ***	(0.0399) **	(0.0432) ***	(0.0397) **	(0.0402) ***	(0.0403) ***
Total Adjusted Gross Income	-0.1848	-0.1493	-0.1323	-0.1317	-0.2033	-0.1757	-0.1715	-0.1711	-0.1533	-0.1475	-0.1477	-0.1456
	(0.0218) ***	(0.0199) ***	(0.0205) ***	(0.0205) ***	(0.0216) ***	(0.0198) ***	(0.0204) ***	(0.0204) ***	(0.0227) ***	(0.0208) ***	(0.0210) ***	(0.0210) ***
Squared(Total Adjusted Gross Income)	0.0702	0.0703	0.0569	0.0586	0.0858	0.0905	0.0852	0.0865	0.0563	0.0719	0.0696	0.0712
	(0.0196) ***	(0.0178) **	(0.0177) *	(0.0178) **	(0.0195) ***	(0.0178) **	(0.0177) **	(0.0177) **	(0.0203) **	(0.0185) **	(0.0182) **	(0.0182) **
Qualifying for 20 Percent Credit	0.1340	0.0974	0.1038	0.1039	0.1555	0.1222	0.1282	0.1278	0.1764	0.1332	0.1500	0.1488
	(0.0086) ***	(0.0078) ***	(0.0076) ***	(0.0076) ***	(0.0086) ***	(0.0079) ***	(0.0076) ***	(0.0077) ***	(0.0088) ***	(0.0081) ***	(0.0079) ***	(0.0079) ***
Qualifying for 10 Percent Credit	0.5016	0.4181	0.4448	0.4423	0.5169	0.4258	0.4521	0.4499	0.4843	0.3937	0.4261	0.4239
	(0.0091) ***	(0.0086) ***	(0.0088) ***	(0.0088) ***	(0.0090) ***	(0.0086) ***	(0.0088) ***	(0.0088) ***	(0.0093) ***	(0.0088) ***	(0.0089) ***	(0.0089) ***
90 percent or more of income from SE [4]	0.0339	0.0438	0.0502	0.0495	0.0408	0.0415	0.0389	0.0372	0.1776	0.1390	0.1031	0.1015
	(0.0084) ***	(0.0077) ***	(0.0092) ***	(0.0092) **	(0.0083) ***	(0.0076) ***	(0.0089) **	(0.0090) **	(0.0106) ***	(0.0097) ***	(0.0103) ***	(0.0104) ***
Male	0.0428	0.0582	0.0683	0.0698	0.0423	0.0551	0.0648	0.0660	0.0717	0.0770	0.0807	0.0835
	(0.0097) ***	(0.0089) ***	(0.0087) ***	(0.0087) ***	(0.0095) ***	(0.0086) ***	(0.0085) ***	(0.0085) ***	(0.0097) ***	(0.0088) ***	(0.0086) ***	(0.0086) ***
Between 25 and 34	-0.0015	-0.0112	-0.0664	-0.0657	0.0590	0.0218	-0.0357	-0.0359	0.0299	0.0269	-0.0595	-0.0612
	(0.0171)	(0.0156)	(0.0157) **	(0.0157) **	(0.0218) **	(0.0199)	(0.0201)	(0.0202)	(0.0260)	(0.0238)	(0.0238) .	(0.0239) .
Between 35 and 44	-0.0804	-0.0523	-0.0706	-0.0688	-0.0340	-0.0284	-0.0544	-0.0531	-0.0338	-0.0170	-0.0387	-0.0367
	(0.0107) ***	(0.0097) **	(0.0098) ***	(0.0098) ***	(0.0118) **	(0.0107) *	(0.0108) **	(0.0108) **	(0.0132) *	(0.0121)	(0.0121) *	(0.0121) *
Between 45 and 54	-0.0470	-0.0464	-0.0934	-0.0930	-0.0028	-0.0243	-0.0680	-0.0678	-0.0146	-0.0170	-0.0590	-0.0598
	(0.0135) ***	(0.0122) **	(0.0124) ***	(0.0124) ***	(0.0153)	(0.0140)	(0.0140) **	(0.0140) **	(0.0173)	(0.0158)	(0.0156) **	(0.0156) **
Between 55 and 59	-0.0652	-0.0778	-0.0956	-0.0948	-0.0018	-0.0317	-0.0531	-0.0532	0.0052	-0.0209	-0.0726	-0.0713
	(0.0137) ***	(0.0124) ***	(0.0123) ***	(0.0123) ***	(0.0147)	(0.0134) .	(0.0132) **	(0.0132) **	(0.0167)	(0.0153)	(0.0150) **	(0.0150) **
Married filing Jointly	0.5571	0.3503	0.3323	0.3244	0.4959	0.2760	0.2485	0.2410	0.4248	0.2581	0.1878	0.1770
	(0.0289) ***	(0.0268) ***	(0.0281) ***	(0.0282) ***	(0.0287) ***	(0.0267) ***	(0.0280) ***	(0.0280) ***	(0.0294) ***	(0.0271) ***	(0.0280) ***	(0.0280) ***
Head of Household	0.3375	0.2527	0.3298	0.3289	0.2322	0.1421	0.2133	0.2132	0.1781	0.1327	0.1594	0.1569
	(0.0353) ***	(0.0322) ***	(0.0346) ***	(0.0346) ***	(0.0349) ***	(0.0320) **	(0.0341) ***	(0.0341) ***	(0.0357) ***	(0.0326) **	(0.0342) **	(0.0342) **
1 to 3 dependents	-0.0966	-0.0730	-0.0981	-0.0960	-0.0695	-0.0351	-0.0473	-0.0455	-0.0267	-0.0097	-0.0379	-0.0340
	(0.0168) ***	(0.0153) **	(0.0158) ***	(0.0158) ***	(0.0166) ***	(0.0152) .	(0.0156) *	(0.0156) *	(0.0169)	(0.0155)	(0.0158) .	(0.0158) .
4 or more dependents	0.0013	0.0188	-0.0159	-0.0139	0.0502	0.0726	0.0538	0.0551	0.1197	0.1019	0.1037	0.1079
	(0.0271)	(0.0247)	(0.0260)	(0.0260)	(0.0264) .	(0.0241) *	(0.0253) .	(0.0253) .	(0.0263) ***	(0.0240) **	(0.0250) **	(0.0250) **
Using a paid preparer	-0.3729	-0.0080	-0.0974	-0.1042	-0.4359	-0.0718	-0.1628	-0.1761	-0.5675	-0.1568	-0.2659	-0.2789
	(0.0581) ***	(0.0539)	(0.0653)	(0.0660)	(0.0572) ***	(0.0533)	(0.0647) .	(0.0653) *	(0.0585) ***	(0.0549) *	(0.0653) **	(0.0659) **
Squared(Using a paid preparer)	0.4377	0.0620	0.2787	0.2858	0.4634	0.1012	0.3157	0.3300	0.6058	0.1937	0.4168	0.4284
	(0.0578) ***	(0.0538)	(0.0627) **	(0.0634) **	(0.0569) ***	(0.0531)	(0.0619) **	(0.0625) **	(0.0580) ***	(0.0546) **	(0.0625) ***	(0.0630) ***
ρ		0.5640		0.7992		0.5481		0.7688		0.5536		0.7446
		(0.0130) ***		(0.0226) ***		(0.0130) ***		(0.0248) ***		(0.0131) ***		(0.0258) ***
λ			0.9292				0.9253				0.9286	
			(0.0117) ***				(0.0121) ***				(0.0118) ***	

Footnotes at the end of table.

Table D.1: Full Population Model Estimates, 2002-2013, cont.

	2005				2006				2007			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Lag(Making Retirement Contributions)				-(0.5899)				-(0.6655)				-(0.6094)
				(0.1696) **				(0.1664) **				(0.1756) **
Lag(Squared(Making Retirement Contributions))				(0.2887)				(0.3479)				(0.3554)
				(0.1738)				(0.1706) .				(0.1700) .
Lag(Total Adjusted Gross Income)				(0.0391)				(0.0981)				(0.2075)
				(0.0868)				(0.0865)				(0.0891) .
Lag(Squared(Total Adjusted Gross Income))				(0.0294)				-(0.0422)				-(0.0779)
				(0.0919)				(0.0921)				(0.0918)
Lag(Qualifying for 20 Percent Credit)				(0.0148)				-(0.0197)				-(0.0978)
				(0.0537)				(0.0529)				(0.0505)
Lag(Qualifying for 10 Percent Credit)				-(0.3162)				-(0.2686)				-(0.2099)
				(0.0430) ***				(0.0440) ***				(0.0484) **
Lag(90 Percent or more of income from SE[4])				(0.0150)				(0.0209)				(0.0895)
				(0.0223)				(0.0231)				(0.0325) *
Lag(Male)				-(0.0564)				-(0.0686)				(0.0263)
				(0.0489)				(0.0525)				(0.0541)
Lag(Between 25 and 34)				(0.1303)				(0.0625)				-(0.0374)
				(0.0791)				(0.0990)				(0.1077)
Lag(Between 35 and 44)				(0.0802)				(0.0753)				(0.0014)
				(0.0477)				(0.0501)				(0.0553)
Lag(Between 45 and 54)				(0.0939)				(0.0645)				-(0.0785)
				(0.0582)				(0.0691)				(0.0775)
Lag(Between 55 and 59)				(0.1825)				(0.0653)				(0.2002)
				(0.0689) *				(0.0787)				(0.0760) *
Lag(Married filing Jointly)				-(0.2228)				-(0.1045)				-(0.1948)
				(0.0888) .				(0.0911)				(0.0946) .
Lag(Head of Household)				-(0.3289)				-(0.1982)				-(0.2232)
				(0.1068) *				(0.1093)				(0.1108) .
Lag(1 to 3 dependents)				(0.1397)				(0.0517)				(0.1850)
				(0.0539) *				(0.0557)				(0.0552) **
Lag(4 or more dependents)				(0.1272)				(0.0038)				(0.1243)
				(0.0825)				(0.0859)				(0.0862)
Lag(Using a paid preparer)				(0.2946)				(0.3147)				(0.5111)
				(0.1488) .				(0.1522) .				(0.1507) **
Lag(Squared(Using a paid preparer))				-(0.4700)				-(0.4681)				-(0.6526)
				(0.1483) *				(0.1509) *				(0.1495) **

Footnotes at the end of table.

Table D.1: Full Population Model Estimates, 2002-2013, cont.

	2005				2006				2007			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
N	7,455	7,455	7,455	7,455	7,453				7,476			
AIC	14,714	13,354	12,955	12,900	14,553	13,251	12,864	12,797	14,983	13,687	13,160	13,070
Moran I Residuals	0.1559 ***	0.0446 ***	-0.0035	-0.0005	0.1472 ***	0.0439 ***	-0.0044	-0.0009	0.1547 ***	0.0468 ***	-0.0065	-0.0024
Log Likelihood		-6,655.98	-6,456.58	-6,411.00		-6,604.39	-6,410.84	-6,359.26		-6,822.71	-6,558.79	-6,496.17
Lagrange Multiplier												
LMerr	6,004 ***				5,355 ***				5,951 ***			
LMlag	2,342 ***				2,179 ***				2,247 ***			
RLMerr	4,244 ***				3,770 ***				4,236 ***			
RLMlag	581 ***				594 ***				532 ***			

Footnotes at the end of table.

Table D.1: Full Population Model Estimates, 2002-2013, cont.

	2008				2009				2010			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Intercept	0.0000	0.0078	0.0900	0.0088	0.0000	0.0077	0.1105	0.0110	0.0000	0.0072	0.0879	0.0103
	(0.0074)	(0.0068)	(0.0809)	(0.0066)	(0.0074)	(0.0068)	(0.0817)	(0.0066)	(0.0076)	(0.0070)	(0.0738)	(0.0068)
Making Retirement Contributions	0.7430	0.5778	0.6554	0.6612	0.8419	0.6628	0.7650	0.7722	0.7659	0.5691	0.6746	0.6787
	(0.0437) ***	(0.0402) ***	(0.0412) ***	(0.0412) ***	(0.0439) ***	(0.0403) ***	(0.0411) ***	(0.0411) ***	(0.0444) ***	(0.0413) ***	(0.0424) ***	(0.0424) ***
Squared(Making Retirement Contributions)	-0.4375	-0.3048	-0.2953	-0.2968	-0.4964	-0.3560	-0.3528	-0.3545	-0.4523	-0.2853	-0.2874	-0.2856
	(0.0436) ***	(0.0399) ***	(0.0406) ***	(0.0406) ***	(0.0438) ***	(0.0400) ***	(0.0406) ***	(0.0406) ***	(0.0443) ***	(0.0409) ***	(0.0418) ***	(0.0418) ***
Total Adjusted Gross Income	-0.1876	-0.1707	-0.1809	-0.1803	-0.1420	-0.1393	-0.1776	-0.1769	-0.1351	-0.1305	-0.1739	-0.1731
	(0.0217) ***	(0.0198) ***	(0.0205) ***	(0.0205) ***	(0.0216) ***	(0.0196) ***	(0.0202) ***	(0.0202) ***	(0.0219) ***	(0.0201) ***	(0.0208) ***	(0.0208) ***
Squared(Total Adjusted Gross Income)	0.0790	0.0886	0.0937	0.0950	0.0476	0.0660	0.0883	0.0894	0.0421	0.0602	0.0842	0.0854
	(0.0196) ***	(0.0179) **	(0.0179) **	(0.0179) **	(0.0194) *	(0.0177) **	(0.0176) **	(0.0176) **	(0.0197) *	(0.0182) **	(0.0182) **	(0.0181) **
Qualifying for 20 Percent Credit	0.1357	0.0919	0.0979	0.0975	0.1335	0.0933	0.0913	0.0901	0.1468	0.1152	0.1084	0.1102
	(0.0084) ***	(0.0078) ***	(0.0076) ***	(0.0076) ***	(0.0085) ***	(0.0078) ***	(0.0076) ***	(0.0077) ***	(0.0084) ***	(0.0077) ***	(0.0076) ***	(0.0076) ***
Qualifying for 10 Percent Credit	0.5143	0.4265	0.4422	0.4380	0.4965	0.4139	0.4188	0.4153	0.4543	0.3833	0.3911	0.3864
	(0.0089) ***	(0.0084) ***	(0.0087) ***	(0.0088) ***	(0.0089) ***	(0.0083) ***	(0.0086) ***	(0.0086) ***	(0.0089) ***	(0.0084) ***	(0.0087) ***	(0.0087) ***
90 percent or more of income from SE [4]	0.0384	0.0337	0.0175	0.0164	0.0416	0.0425	0.0202	0.0202	0.0154	0.0257	0.0124	0.0127
	(0.0084) ***	(0.0077) **	(0.0091)	(0.0092)	(0.0086) ***	(0.0078) ***	(0.0092) .	(0.0092) .	(0.0088) .	(0.0081) *	(0.0095)	(0.0095)
Male	0.0323	0.0360	0.0420	0.0434	0.0842	0.0803	0.0849	0.0865	0.0639	0.0589	0.0642	0.0653
	(0.0098) ***	(0.0089) **	(0.0088) **	(0.0088) **	(0.0099) ***	(0.0090) ***	(0.0088) ***	(0.0088) ***	(0.0101) ***	(0.0093) ***	(0.0092) ***	(0.0092) ***
Between 25 and 34	0.1047	0.0966	0.0736	0.0726	0.0455	0.0542	0.0455	0.0424	0.0732	0.0811	0.0700	0.0676
	(0.0246) ***	(0.0224) **	(0.0231) *	(0.0231) *	(0.0206) *	(0.0188) *	(0.0194) .	(0.0194) .	(0.0180) ***	(0.0166) **	(0.0173) **	(0.0173) **
Between 35 and 44	0.0482	0.0510	0.0204	0.0228	0.0388	0.0445	0.0162	0.0174	0.0497	0.0506	0.0173	0.0188
	(0.0133) ***	(0.0121) **	(0.0125)	(0.0125)	(0.0120) **	(0.0110) **	(0.0114)	(0.0114)	(0.0120) ***	(0.0110) **	(0.0116)	(0.0116)
Between 45 and 54	0.0281	0.0266	0.0020	0.0022	0.0694	0.0666	0.0579	0.0575	0.0592	0.0623	0.0488	0.0474
	(0.0162) .	(0.0148)	(0.0149)	(0.0149)	(0.0141) ***	(0.0129) **	(0.0130) **	(0.0130) **	(0.0131) ***	(0.0120) **	(0.0122) **	(0.0122) **
Between 55 and 59	0.0670	0.0394	0.0208	0.0213	0.0121	-0.0031	-0.0137	-0.0145	0.0680	0.0430	0.0300	0.0297
	(0.0160) ***	(0.0146) *	(0.0145)	(0.0145)	(0.0146)	(0.0132)	(0.0131)	(0.0131)	(0.0139) ***	(0.0128) **	(0.0127) .	(0.0127) .
Married filing Jointly	0.4460	0.2192	0.2064	0.1944	0.5020	0.2793	0.2438	0.2305	0.4352	0.2192	0.2020	0.1897
	(0.0298) ***	(0.0277) ***	(0.0289) ***	(0.0289) ***	(0.0294) ***	(0.0272) ***	(0.0281) ***	(0.0281) ***	(0.0290) ***	(0.0272) ***	(0.0280) ***	(0.0280) ***
Head of Household	0.1937	0.0976	0.1630	0.1582	0.2507	0.1688	0.1890	0.1829	0.1472	0.0724	0.1132	0.1101
	(0.0349) ***	(0.0320) *	(0.0339) **	(0.0340) **	(0.0337) ***	(0.0308) ***	(0.0324) ***	(0.0323) **	(0.0335) ***	(0.0309) .	(0.0322) **	(0.0321) **
1 to 3 dependents	-0.0731	-0.0307	-0.0379	-0.0339	-0.1380	-0.0797	-0.0744	-0.0684	-0.0724	-0.0182	-0.0138	-0.0079
	(0.0170) ***	(0.0156) .	(0.0160) .	(0.0160) .	(0.0168) ***	(0.0153) **	(0.0156) **	(0.0156) **	(0.0169) ***	(0.0156)	(0.0160)	(0.0160)
4 or more dependents	0.0895	0.0955	0.0892	0.0931	0.0483	0.0529	0.1001	0.1066	0.1190	0.1134	0.1657	0.1695
	(0.0263) ***	(0.0240) **	(0.0256) **	(0.0256) **	(0.0253) .	(0.0230) .	(0.0245) **	(0.0245) **	(0.0253) ***	(0.0233) **	(0.0245) ***	(0.0245) ***
Using a paid preparer	-0.4506	-0.0932	-0.1089	-0.1152	-0.5296	-0.2198	-0.2720	-0.2896	-0.4942	-0.2235	-0.3101	-0.3241
	(0.0578) ***	(0.0537)	(0.0659)	(0.0665)	(0.0581) ***	(0.0536) **	(0.0652) **	(0.0657) **	(0.0591) ***	(0.0550) **	(0.0671) **	(0.0676) **
Squared(Using a paid preparer)	0.4859	0.1325	0.2581	0.2643	0.5732	0.2545	0.3856	0.3985	0.5202	0.2442	0.3833	0.3946
	(0.0573) ***	(0.0533) .	(0.0623) **	(0.0629) **	(0.0576) ***	(0.0533) **	(0.0614) ***	(0.0618) ***	(0.0585) ***	(0.0545) **	(0.0628) ***	(0.0632) ***
ρ		0.5427		0.7496		0.5405		0.7502		0.5388		0.7156
		(0.0130) ***		(0.0258) ***		(0.0128) ***		(0.0254) ***		(0.0136) ***		(0.0278) ***
λ			0.9185				0.9199				0.9076	
			(0.0129) ***				(0.0127) ***				(0.0140) ***	

Footnotes at the end of table.

Table D.1: Full Population Model Estimates, 2002-2013, cont.

	2008				2009				2010			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Lag(Making Retirement Contributions)				-(0.5004)				-(0.6912)				-(0.4589)
				(0.1808) *				(0.1761) **				(0.1798) .
Lag(Squared(Making Retirement Contributions))				(0.2590)				(0.3988)				(0.2326)
				(0.1830)				(0.1765) .				(0.1812)
Lag(Total Adjusted Gross Income)				(0.1692)				(0.2275)				(0.2199)
				(0.0879)				(0.0868) *				(0.0893) .
Lag(Squared(Total Adjusted Gross Income))				-(0.1300)				-(0.1669)				-(0.1605)
				(0.0912)				(0.0890)				(0.0912)
Lag(Qualifying for 20 Percent Credit)				(0.0054)				(0.0112)				(0.1701)
				(0.0594)				(0.0520)				(0.0560) *
Lag(Qualifying for 10 Percent Credit)				-(0.2756)				-(0.2587)				-(0.3273)
				(0.0437) ***				(0.0424) ***				(0.0409) ***
Lag(90 Percent or more of income from SE[4])				(0.0592)				(0.0649)				(0.0701)
				(0.0225) *				(0.0233) *				(0.0236) *
Lag(Male)				(0.0087)				(0.0156)				(0.0506)
				(0.0504)				(0.0529)				(0.0516)
Lag(Between 25 and 34)				-(0.0654)				-(0.2144)				-(0.1408)
				(0.1086)				(0.0874) .				(0.0772)
Lag(Between 35 and 44)				(0.0723)				-(0.0162)				(0.0419)
				(0.0544)				(0.0467)				(0.0444)
Lag(Between 45 and 54)				-(0.0289)				-(0.1385)				-(0.1030)
				(0.0733)				(0.0657) .				(0.0606)
Lag(Between 55 and 59)				(0.1002)				-(0.0333)				(0.0559)
				(0.0859)				(0.0713)				(0.0645)
Lag(Married filing Jointly)				-(0.1751)				-(0.1222)				-(0.0192)
				(0.0957)				(0.0998)				(0.0994)
Lag(Head of Household)				-(0.2552)				-(0.1747)				-(0.0569)
				(0.1114) .				(0.1114)				(0.1150)
Lag(1 to 3 dependents)				(0.0964)				(0.1015)				(0.0366)
				(0.0598)				(0.0607)				(0.0605)
Lag(4 or more dependents)				(0.0651)				(0.0003)				-(0.1114)
				(0.0850)				(0.0821)				(0.0850)
Lag(Using a paid preparer)				(0.2659)				(0.3693)				(0.4907)
				(0.1485)				(0.1491) .				(0.1523) *
Lag(Squared(Using a paid preparer))				-(0.4093)				-(0.4647)				-(0.5684)
				(0.1476) *				(0.1486) *				(0.1520) **

Footnotes at the end of table.

Table D.1: Full Population Model Estimates, 2002-2013, cont.

	2008				2009				2010			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
N	7,449				7,438				7,424			
AIC	14,573	13,262	12,949	12,871	14,505	13,173	12,822	12,732	14,879	13,700	13,396	13,294
Moran I Residuals	0.1480 ***	0.0402 ***	-0.0046	-0.0011	0.1541 ***	0.0434 ***	-0.0035	-0.0003	0.1412 ***	0.0407 ***	-0.0041	-0.0005
Log Likelihood		-6,610.13	-6,453.47	-6,396.40		-6,565.54	-6,389.98	-6,326.77		-6,828.98	-6,676.94	-6,608.15
Lagrange Multiplier												
LMerr	5,409 ***				5,846 ***				4,890 ***			
LMlag	2,213 ***				2,280 ***				1,992 ***			
RLMerr	3,812 ***				4,175 ***				3,413 ***			
RLMlag	616 ***				609 ***				515 ***			

Footnotes at the end of table.

Table D.1: Full Population Model Estimates, 2002-2013, cont.

	2011				2012				2013			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Intercept	0.0000	0.0057	0.0660	0.0088	0.0000	0.0056	0.0795	0.0105	0.0000	0.0047	0.0641	0.0093
	(0.0074)	(0.0070)	(0.0641)	(0.0067)	(0.0077)	(0.0072)	(0.0653)	(0.0069)	(0.0077)	(0.0073)	(0.0644)	(0.0070)
Making Retirement Contributions	0.7138	0.5538	0.6896	0.6956	0.5709	0.3978	0.5386	0.5471	0.6836	0.5209	0.7036	0.7130
	(0.0448) ***	(0.0425) ***	(0.0431) ***	(0.0432) ***	(0.0461) ***	(0.0437) ***	(0.0445) ***	(0.0446) ***	(0.0441) ***	(0.0421) ***	(0.0427) ***	(0.0427) ***
Squared(Making Retirement Contributions)	-0.3877	-0.2588	-0.2841	-0.2827	-0.2353	-0.0979	-0.1364	-0.1387	-0.3521	-0.2177	-0.2885	-0.2910
	(0.0446) ***	(0.0420) ***	(0.0424) ***	(0.0425) ***	(0.0460) ***	(0.0433) .	(0.0439) *	(0.0439) *	(0.0440) ***	(0.0416) **	(0.0418) ***	(0.0418) ***
Total Adjusted Gross Income	-0.1302	-0.1339	-0.1642	-0.1636	-0.1100	-0.1078	-0.1550	-0.1534	-0.0878	-0.1029	-0.1468	-0.1460
	(0.0212) ***	(0.0199) ***	(0.0203) ***	(0.0203) ***	(0.0209) ***	(0.0196) ***	(0.0200) ***	(0.0200) ***	(0.0219) ***	(0.0206) **	(0.0210) ***	(0.0210) ***
Squared(Total Adjusted Gross Income)	0.0423	0.0636	0.0758	0.0772	0.0256	0.0406	0.0656	0.0655	0.0207	0.0506	0.0746	0.0762
	(0.0191) *	(0.0179) **	(0.0177) **	(0.0177) **	(0.0186)	(0.0174) .	(0.0172) **	(0.0172) **	(0.0197)	(0.0185) *	(0.0184) **	(0.0183) **
Qualifying for 20 Percent Credit	0.1408	0.1128	0.1116	0.1113	0.1174	0.0909	0.0854	0.0841	0.1161	0.0887	0.0808	0.0797
	(0.0082) ***	(0.0077) ***	(0.0075) ***	(0.0075) ***	(0.0086) ***	(0.0081) ***	(0.0079) ***	(0.0079) ***	(0.0086) ***	(0.0081) ***	(0.0079) ***	(0.0079) ***
Qualifying for 10 Percent Credit	0.4919	0.4293	0.4360	0.4336	0.4992	0.4330	0.4252	0.4219	0.4927	0.4300	0.4204	0.4173
	(0.0086) ***	(0.0083) ***	(0.0086) ***	(0.0087) ***	(0.0090) ***	(0.0087) ***	(0.0090) ***	(0.0090) ***	(0.0092) ***	(0.0088) ***	(0.0091) ***	(0.0091) ***
90 percent or more of income from SE [4]	0.0159	0.0194	0.0055	0.0064	0.0227	0.0191	-0.0001	0.0001	0.0331	0.0337	0.0257	0.0259
	(0.0086) .	(0.0080) .	(0.0095)	(0.0095)	(0.0088) *	(0.0083) .	(0.0098)	(0.0098)	(0.0089) ***	(0.0084) **	(0.0098) *	(0.0098) *
Male	0.0513	0.0481	0.0505	0.0522	0.0345	0.0347	0.0348	0.0362	0.0338	0.0325	0.0378	0.0382
	(0.0100) ***	(0.0093) **	(0.0092) ***	(0.0092) **	(0.0106) **	(0.0099) **	(0.0098) **	(0.0098) **	(0.0108) **	(0.0101) *	(0.0100) **	(0.0100) **
Between 25 and 34	-0.0008	0.0138	0.0183	0.0151	0.0201	0.0445	0.0477	0.0453	0.0066	0.0359	0.0357	0.0334
	(0.0157)	(0.0147)	(0.0151)	(0.0151)	(0.0149)	(0.0140) *	(0.0144) **	(0.0144) *	(0.0139)	(0.0131) *	(0.0133) *	(0.0133) .
Between 35 and 44	0.0282	0.0261	-0.0135	-0.0122	0.0646	0.0671	0.0327	0.0328	0.0650	0.0637	0.0138	0.0142
	(0.0114) *	(0.0107) .	(0.0112)	(0.0112)	(0.0118) ***	(0.0111) ***	(0.0116) *	(0.0116) *	(0.0121) ***	(0.0114) ***	(0.0122)	(0.0122)
Between 45 and 54	0.0332	0.0353	0.0267	0.0261	0.0467	0.0476	0.0278	0.0278	0.0639	0.0589	0.0358	0.0364
	(0.0115) **	(0.0107) *	(0.0108) .	(0.0108) .	(0.0109) ***	(0.0102) **	(0.0104) *	(0.0104) *	(0.0101) ***	(0.0095) ***	(0.0097) **	(0.0097) **
Between 55 and 59	0.0049	-0.0095	-0.0196	-0.0204	0.0325	0.0275	0.0236	0.0223	0.0321	0.0301	0.0176	0.0164
	(0.0126)	(0.0118)	(0.0117)	(0.0117)	(0.0124) **	(0.0117) .	(0.0115) .	(0.0115)	(0.0118) **	(0.0111) *	(0.0109)	(0.0109)
Married filing Jointly	0.3652	0.1635	0.1783	0.1676	0.4224	0.2497	0.2260	0.2154	0.4064	0.2500	0.1871	0.1784
	(0.0269) ***	(0.0260) ***	(0.0269) ***	(0.0269) ***	(0.0290) ***	(0.0276) ***	(0.0281) ***	(0.0281) ***	(0.0289) ***	(0.0273) ***	(0.0276) ***	(0.0275) ***
Head of Household	0.0678	-0.0176	0.0519	0.0507	0.2186	0.1519	0.1771	0.1766	0.1754	0.1279	0.1138	0.1150
	(0.0312) *	(0.0294)	(0.0310)	(0.0310)	(0.0330) ***	(0.0310) **	(0.0324) ***	(0.0324) **	(0.0331) ***	(0.0311) **	(0.0322) **	(0.0321) **
1 to 3 dependents	-0.0696	-0.0083	-0.0214	-0.0177	-0.0871	-0.0438	-0.0295	-0.0251	-0.1059	-0.0626	-0.0247	-0.0215
	(0.0158) ***	(0.0150)	(0.0153)	(0.0153)	(0.0168) ***	(0.0157) *	(0.0160)	(0.0160)	(0.0167) ***	(0.0157) **	(0.0159)	(0.0159)
4 or more dependents	0.1500	0.1593	0.1873	0.1893	0.0862	0.0710	0.1512	0.1545	0.1225	0.0992	0.2164	0.2179
	(0.0238) ***	(0.0223) ***	(0.0233) ***	(0.0233) ***	(0.0247) ***	(0.0232) *	(0.0247) ***	(0.0247) ***	(0.0250) ***	(0.0236) **	(0.0250) ***	(0.0250) ***
Using a paid preparer	-0.4912	-0.2714	-0.3424	-0.3541	-0.3765	-0.1395	-0.2811	-0.3018	-0.4788	-0.2684	-0.4760	-0.5015
	(0.0569) ***	(0.0539) **	(0.0646) **	(0.0652) **	(0.0586) ***	(0.0555) .	(0.0669) **	(0.0676) **	(0.0588) ***	(0.0559) **	(0.0669) ***	(0.0675) ***
Squared(Using a paid preparer)	0.5235	0.2991	0.4236	0.4320	0.4183	0.1761	0.3668	0.3831	0.5393	0.3143	0.5569	0.5774
	(0.0563) ***	(0.0534) ***	(0.0607) ***	(0.0612) ***	(0.0580) ***	(0.0551) *	(0.0631) ***	(0.0637) ***	(0.0583) ***	(0.0556) ***	(0.0632) ***	(0.0637) ***
ρ		0.4777		0.7220		0.4997		-0.6072		0.5023		-0.7875
		(0.0141) ***		(0.0277) ***		(0.0145) ***		(0.6999)		(0.0148) ***		(0.6230)
λ			0.8952				0.8939	0.0289 ***			0.8911	0.0333 ***
			(0.0152) ***				(0.0154) ***				(0.0156) ***	

Footnotes at the end of table.

Table D.1: Full Population Model Estimates, 2002-2013, cont.

	2011				2012				2013			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Lag(Making Retirement Contributions)				-(0.7014)								
				(0.1827) **				(0.1884) *				(0.1998) **
Lag(Squared(Making Retirement Contributions))				(0.4251)				(0.3063)				(0.4802)
				(0.1853) .				(0.1922)				(0.2053) .
Lag(Total Adjusted Gross Income)				(0.1119)				(0.2326)				(0.1831)
				(0.0916)				(0.0913) .				(0.0947)
Lag(Squared(Total Adjusted Gross Income))				-(0.0648)				-(0.1946)				-(0.1346)
				(0.0929)				(0.0938) .				(0.0964)
Lag(Qualifying for 20 Percent Credit)				-(0.0030)				-(0.0136)				(0.0524)
				(0.0580)				(0.0626)				(0.0572)
Lag(Qualifying for 10 Percent Credit)				-(0.2776)				-(0.1906)				-(0.1159)
				(0.0414) ***				(0.0453) **				(0.0483) .
Lag(90 Percent or more of income from SE[4])				(0.0748)				(0.0620)				(0.0152)
				(0.0245) *				(0.0261) .				(0.0266)
Lag(Male)				(0.0528)				(0.0360)				-(0.0151)
				(0.0515)				(0.0534)				(0.0539)
Lag(Between 25 and 34)				-(0.2009)				-(0.1509)				-(0.1439)
				(0.0701) *				(0.0673) .				(0.0627) .
Lag(Between 35 and 44)				(0.0814)				(0.0239)				(0.0693)
				(0.0444)				(0.0449)				(0.0435)
Lag(Between 45 and 54)				-(0.1033)				(0.0176)				(0.0783)
				(0.0541)				(0.0518)				(0.0472)
Lag(Between 55 and 59)				-(0.0136)				-(0.0660)				-(0.0693)
				(0.0616)				(0.0619)				(0.0600)
Lag(Married filing Jointly)				-(0.0248)				-(0.0811)				(0.1002)
				(0.0874)				(0.1006)				(0.1099)
Lag(Head of Household)				(0.0210)				-(0.1069)				(0.0397)
				(0.1060)				(0.1145)				(0.1223)
Lag(1 to 3 dependents)				(0.0056)				(0.0181)				-(0.1069)
				(0.0535)				(0.0602)				(0.0647)
Lag(4 or more dependents)				-(0.1541)				-(0.1173)				-(0.2580)
				(0.0811)				(0.0844)				(0.0890) *
Lag(Using a paid preparer)				(0.4915)				(0.5245)				(0.7949)
				(0.1452) **				(0.1512) **				(0.1512) **
Lag(Squared(Using a paid preparer))				-(0.5640)				-(0.5940)				-(0.8552)
				(0.1444) **				(0.1505) **				(0.1511) **

Footnotes at the end of table.

Table D.1: Full Population Model Estimates, 2002-2013, cont.

	2011				2012				2013			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
N	7,397				7,373	7,373	7,373	7,373	7,355	7,355	7,355	7,355
AIC	14,837	13,487	13,080	13,001	14,762	13,855	13,477	13,383	14,838	13,967	13,590	13,464
Moran I Residuals	0.1319 ***	0.0497 ***	-0.0033	-0.0003	0.1303 ***	0.0451 ***	-0.0034	0.0000	0.1242 ***	0.0428 ***	-0.0046	-0.0005
Log Likelihood		-6,722.31	-6,518.92	-6,461.59		-6,906.50	-6,717.64	-6,652.57		-6,962.45	-6,773.97	-6,692.82
Lagrange Multiplier												
LMerr	4,237 ***				4,109 ***				3,711 ***			
LMlag	1,421 ***				1,505 ***				1,454 ***			
RLMerr	3,134 ***				2,926 ***				2,577 ***			
RLMlag	318 ***				322 ***				320 ***			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Standard Errors in Paretheses below parameter estimates
[1] Population reflects tax filers between 18 and 57 in 1999 with income below $100,000.
[2] All parameters reflect the percent for a given ZIP Code except Total AGI, which is the sum of reported AGI
[3] Lagged parameters are represented by θ in the SDM spatial model.
[4] SE - Self-Employed
Source: author's computation

Table D.2: Subgroup 1 Model Estimates, 2002-2013

	2002				2003				2004			
Parameter [1][2]	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Intercept	0.0161	0.0043	0.0239	0.0041	-0.0052	0.0032	0.0164	0.0030	-0.0654	0.0031	0.0176	0.0038
	(0.0329)	(0.0071)	(0.0488)	(0.0070)	(0.0313)	(0.0066)	(0.0430)	(0.0065)	(0.0316) *	(0.0065)	(0.0376)	(0.0064)
Making Retirement Contributions	0.7295	0.6226	0.6077	0.6059	0.8094	0.7062	0.6929	0.6855	0.8891	0.7758	0.7838	0.7772
	(0.0385) ***	(0.0364) ***	(0.0367) ***	(0.0367) ***	(0.0327) ***	(0.0312) ***	(0.0314) ***	(0.0314) ***	(0.0281) ***	(0.0272) ***	(0.0272) ***	(0.0272) ***
Squared(Making Retirement Contributions)	-0.0744	-0.0634	-0.0231	-0.0251	-0.1173	-0.0922	-0.0604	-0.0564	-0.1994	-0.1565	-0.1514	-0.1472
	(0.0374) *	(0.0352)	(0.0354)	(0.0354)	(0.0317) ***	(0.0301) *	(0.0301)	(0.0301)	(0.0272) ***	(0.0259) ***	(0.0258) **	(0.0259) **
Total Adjusted Gross Income	0.0308	0.0104	0.0029	0.0016	0.0626	0.0320	0.0297	0.0277	0.0564	0.0333	0.0337	0.0322
	(0.0109) **	(0.0102)	(0.0108)	(0.0108)	(0.0195) **	(0.0187)	(0.0193)	(0.0194)	(0.0190) **	(0.0182)	(0.0190)	(0.0190)
Squared(Total Adjusted Gross Income)	-0.0400	-0.0281	-0.0279	-0.0291	-0.0502	-0.0276	-0.0274	-0.0255	-0.0528	-0.0347	-0.0364	-0.0349
	(0.0098) ***	(0.0092) *	(0.0091) *	(0.0092) *	(0.0182) **	(0.0174)	(0.0176)	(0.0176)	(0.0178) **	(0.0171) .	(0.0173) .	(0.0173) .
Qualifying for 20 Percent Credit	-0.0234	-0.0220	-0.0203	-0.0195	-0.0007	-0.0003	0.0031	0.0031	-0.0101	-0.0082	-0.0050	-0.0052
	(0.0079) **	(0.0075) *	(0.0074) *	(0.0074) *	(0.0074)	(0.0071)	(0.0070)	(0.0070)	(0.0073)	(0.0070)	(0.0069)	(0.0069)
Qualifying for 10 Percent Credit	0.0054	0.0350	0.0426	0.0460	-0.0274	-0.0013	0.0116	0.0169	-0.0663	-0.0379	-0.0242	-0.0196
	(0.0098)	(0.0093) **	(0.0106) **	(0.0107) **	(0.0091) **	(0.0087)	(0.0099)	(0.0100)	(0.0088) ***	(0.0085) **	(0.0095) .	(0.0096) .
90 percent or more of income from SE [3]	-0.0369	-0.0113	-0.0354	-0.0344	-0.0500	-0.0212	-0.0407	-0.0382	-0.0570	-0.0272	-0.0448	-0.0413
	(0.0087) ***	(0.0083)	(0.0090) **	(0.0091) **	(0.0080) ***	(0.0078) *	(0.0085) **	(0.0085) **	(0.0079) ***	(0.0077) **	(0.0085) **	(0.0086) **
Male	0.0233	0.0201	0.0188	0.0186	0.0280	0.0234	0.0164	0.0160	0.0246	0.0223	0.0170	0.0152
	(0.0084) **	(0.0080) .	(0.0081) .	(0.0081) .	(0.0078) ***	(0.0075) *	(0.0076) .	(0.0076) .	(0.0077) **	(0.0073) *	(0.0075) .	(0.0075) .
Between 25 and 34	-0.0231	-0.0253	-0.0261	-0.0263	-0.0032	-0.0054	-0.0074	-0.0089	-0.0202	-0.0247	-0.0290	-0.0305
	(0.0104) *	(0.0098) .	(0.0101) *	(0.0101) *	(0.0103)	(0.0098)	(0.0100)	(0.0100)	(0.0111) .	(0.0106) .	(0.0108) *	(0.0109) *
Between 35 and 44	-0.0124	-0.0179	-0.0274	-0.0287	0.0116	0.0035	0.0012	-0.0012	0.0130	0.0055	-0.0019	-0.0037
	(0.0103)	(0.0097)	(0.0099) *	(0.0099) *	(0.0095)	(0.0091)	(0.0092)	(0.0092)	(0.0093)	(0.0089)	(0.0090)	(0.0091)
Between 45 and 54	-0.0462	-0.0460	-0.0590	-0.0598	-0.0303	-0.0358	-0.0440	-0.0472	-0.0245	-0.0308	-0.0462	-0.0497
	(0.0114) ***	(0.0107) **	(0.0113) **	(0.0113) **	(0.0104) **	(0.0099) **	(0.0104) **	(0.0105) **	(0.0103) *	(0.0099) *	(0.0103) **	(0.0104) **
Between 55 and 59	-0.0475	-0.0485	-0.0560	-0.0581	-0.0409	-0.0446	-0.0532	-0.0547	-0.0677	-0.0719	-0.0802	-0.0823
	(0.0106) ***	(0.0100) **	(0.0101) **	(0.0101) **	(0.0101) ***	(0.0097) **	(0.0098) **	(0.0098) **	(0.0100) ***	(0.0095) ***	(0.0096) ***	(0.0097) ***
Married filing Jointly	0.4835	0.4067	0.3833	0.3772	0.5006	0.4427	0.4005	0.3916	0.5417	0.4690	0.4431	0.4321
	(0.0254) ***	(0.0241) ***	(0.0254) ***	(0.0255) ***	(0.0229) ***	(0.0221) ***	(0.0233) ***	(0.0235) ***	(0.0220) ***	(0.0213) ***	(0.0223) ***	(0.0224) ***
Head of Household	0.0629	0.1104	0.0962	0.1030	0.0114	0.0569	0.0289	0.0338	0.0720	0.0933	0.0834	0.0859
	(0.0305) *	(0.0289) **	(0.0305) *	(0.0306) **	(0.0280)	(0.0269) .	(0.0283)	(0.0284)	(0.0274) **	(0.0263) **	(0.0278) *	(0.0279) *
1 to 3 dependents	-0.0118	0.0099	0.0132	0.0140	-0.0265	-0.0166	-0.0208	-0.0209	0.0202	0.0280	0.0232	0.0229
	(0.0131)	(0.0124)	(0.0127)	(0.0127)	(0.0121) *	(0.0116)	(0.0119)	(0.0120)	(0.0118) .	(0.0113) .	(0.0116) .	(0.0116) .
4 or more dependents	-0.0454	-0.0308	-0.0511	-0.0531	-0.0258	-0.0178	-0.0491	-0.0526	-0.0344	-0.0172	-0.0453	-0.0483
	(0.0211) *	(0.0199)	(0.0206) .	(0.0206) .	(0.0195)	(0.0186)	(0.0193) .	(0.0194) *	(0.0191) .	(0.0184)	(0.0192) .	(0.0192) .
Using a paid preparer	-0.0758	0.1002	0.1870	0.1919	-0.0697	-0.0026	0.0654	0.0809	0.2136	-0.1174	-0.0480	-0.0269
	(0.2322)	(0.0484) .	(0.0566) **	(0.0574) **	(0.2336)	(0.0478)	(0.0576)	(0.0585)	(0.2354)	(0.0472) .	(0.0565)	(0.0577)
Squared(Using a paid preparer)	0.1262	-0.0921	-0.0633	-0.0601	0.1223	0.0298	0.0888	0.0848	-0.1901	0.1522	0.1964	0.1869
	(0.2526)	(0.0481)	(0.0549)	(0.0555)	(0.2509)	(0.0478)	(0.0558)	(0.0567)	(0.2526)	(0.0474) *	(0.0547) **	(0.0556) **
ρ		0.4703		0.7449		0.3882		0.7078		0.3759		0.6596
		(0.0154) ***		(0.0267) ***		(0.0150) ***		(0.0294) ***		(0.0148) ***		(0.0327) ***
λ			0.8571				0.8486				0.8290	
			(0.0187) ***				(0.0194) ***				(0.0211) ***	

Footnotes at the end of table.

Table D.2: Subgroup 1 Model Estimates, 2002-2013, cont.

	2002				2003				2004			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Lag(Making Retirement Contributions)				(0.1391)				-(0.1901)				-(0.3608)
				(0.1865)				(0.1608)				(0.1613) .
Lag(Squared(Making Retirement Contributions))				-(0.4253)				-(0.1227)				(0.1006)
				(0.1790) .				(0.1570)				(0.1629)
Lag(Total Adjusted Gross Income)				(0.0441)				(0.0222)				-(0.0186)
				(0.0425)				(0.0854)				(0.0830)
Lag(Squared(Total Adjusted Gross Income))				-(0.0947)				(0.0100)				(0.0357)
				(0.0623)				(0.0905)				(0.0895)
Lag(Qualifying for 20 Percent Credit)				(0.0333)				-(0.0505)				-(0.0698)
				(0.0567)				(0.0546)				(0.0551)
Lag(Qualifying for 10 Percent Credit)				-(0.1242)				-(0.0784)				-(0.0894)
				(0.0366) **				(0.0344) .				(0.0354) .
Lag(90 Percent or more of income from SE[3])				(0.0450)				(0.0466)				(0.0449)
				(0.0278)				(0.0264)				(0.0243)
Lag(Male)				-(0.0219)				(0.0304)				-(0.0331)
				(0.0416)				(0.0400)				(0.0383)
Lag(Between 25 and 34)				(0.0152)				-(0.0095)				(0.0124)
				(0.0529)				(0.0543)				(0.0604)
Lag(Between 35 and 44)				(0.0164)				-(0.0402)				(0.0200)
				(0.0497)				(0.0479)				(0.0476)
Lag(Between 45 and 54)				(0.0641)				-(0.0019)				(0.0191)
				(0.0504)				(0.0492)				(0.0534)
Lag(Between 55 and 59)				-(0.0527)				(0.0656)				(0.0698)
				(0.0583)				(0.0584)				(0.0585)
Lag(Married filing Jointly)				-(0.1355)				-(0.1863)				-(0.1857)
				(0.1046)				(0.0964)				(0.0939) .
Lag(Head of Household)				-(0.1044)				-(0.0719)				-(0.0923)
				(0.1183)				(0.1110)				(0.1080)
Lag(1 to 3 dependents)				-(0.1050)				(0.0240)				-(0.0205)
				(0.0546)				(0.0519)				(0.0500)
Lag(4 or more dependents)				(0.0241)				(0.1332)				(0.0717)
				(0.0812)				(0.0786)				(0.0773)
Lag(Using a paid preparer)				-(0.0516)				(0.1304)				(0.0577)
				(0.1359)				(0.1312)				(0.1324)
Lag(Squared(Using a paid preparer))				-(0.0776)				-(0.2949)				-(0.2236)
				(0.1354)				(0.1309) .				(0.1323)

Footnotes at the end of table.

Table D.2: Subgroup 1 Model Estimates, 2002-2013, cont.

	2002				2003				2004			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
N	7,469	7,469	7,469	7,469	7,468	7,468	7,468	7,468	7,462	7,462	7,462	7,462
AIC	14,799	14,012	13,821	13,787	13,567	12,971	12,780	12,739	13,283	12,688	12,556	12,508
Moran I Residuals	0.1187 ***	0.0325 ***	-0.0037	-0.0008	0.0960 ***	0.0326 ***	-0.0018	0.0012	0.0920 ***	0.0250 ***	-0.0051	-0.0023
Log Likelihood		-6,984.98	-6,889.69	-6,854.69		-6,464.75	-6,369.24	-6,330.72		-6,322.80	-6,256.96	-6,215.22
Lagrange Multiplier												
LMerr	3,496 ***				2,286 ***				2,095 ***			
LMlag	1,414 ***				943 ***				963 ***			
RLMerr	2,395 ***				1,619 ***				1,454 ***			
RLMlag	313 ***				276 ***				323 ***			

Footnotes at the end of table.

Table D.2: Subgroup 1 Model Estimates, 2002-2013, cont.

	2005				2006				2007			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Intercept	-0.0651	0.0033	0.0204	0.0039	-0.1338	0.0027	0.0130	0.0031	0.0206	0.0021	0.0121	0.0040
	(0.0292) *	(0.0065)	(0.0353)	(0.0064)	(0.0308) ***	(0.0064)	(0.0351)	(0.0063)	(0.0285)	(0.0064)	(0.0261)	(0.0063)
Making Retirement Contributions	0.6581	0.5725	0.5864	0.5811	0.6816	0.5889	0.6099	0.6050	0.7974	0.6996	0.7278	0.7207
	(0.0313) ***	(0.0302) ***	(0.0303) ***	(0.0303) ***	(0.0278) ***	(0.0270) ***	(0.0270) ***	(0.0270) ***	(0.0262) ***	(0.0259) ***	(0.0259) ***	(0.0259) ***
Squared(Making Retirement Contributions)	0.0413	0.0620	0.0635	0.0660	0.0314	0.0612	0.0619	0.0639	-0.0789	-0.0358	-0.0461	-0.0427
	(0.0305)	(0.0293) .	(0.0293) .	(0.0293) .	(0.0272)	(0.0262) .	(0.0261) .	(0.0261) .	(0.0256) **	(0.0249)	(0.0249)	(0.0249)
Total Adjusted Gross Income	0.0480	0.0320	0.0476	0.0468	0.0071	-0.0043	-0.0003	-0.0019	-0.0121	-0.0245	-0.0083	-0.0106
	(0.0189) *	(0.0182)	(0.0189) .	(0.0190) .	(0.0187)	(0.0180)	(0.0187)	(0.0187)	(0.0185)	(0.0179)	(0.0187)	(0.0188)
Squared(Total Adjusted Gross Income)	-0.0432	-0.0331	-0.0436	-0.0425	-0.0152	-0.0094	-0.0136	-0.0122	-0.0005	0.0043	-0.0077	-0.0055
	(0.0177) *	(0.0171)	(0.0173) .	(0.0173) .	(0.0176)	(0.0169)	(0.0171)	(0.0171)	(0.0173)	(0.0168)	(0.0170)	(0.0170)
Qualifying for 20 Percent Credit	-0.0082	-0.0071	-0.0039	-0.0029	0.0090	0.0108	0.0130	0.0133	-0.0089	-0.0068	-0.0034	-0.0023
	(0.0073)	(0.0070)	(0.0070)	(0.0070)	(0.0073)	(0.0070)	(0.0069)	(0.0070)	(0.0073)	(0.0071)	(0.0071)	(0.0071)
Qualifying for 10 Percent Credit	-0.0268	-0.0033	0.0091	0.0151	-0.0094	0.0050	0.0098	0.0135	-0.0109	0.0035	0.0115	0.0163
	(0.0089) **	(0.0086)	(0.0095)	(0.0097)	(0.0087)	(0.0084)	(0.0092)	(0.0093)	(0.0087)	(0.0084)	(0.0092)	(0.0093)
90 percent or more of income from SE [3]	-0.0533	-0.0300	-0.0432	-0.0398	-0.0514	-0.0269	-0.0429	-0.0405	-0.0634	-0.0475	-0.0649	-0.0624
	(0.0077) ***	(0.0075) **	(0.0081) **	(0.0082) **	(0.0077) ***	(0.0075) **	(0.0082) **	(0.0083) **	(0.0076) ***	(0.0075) ***	(0.0081) ***	(0.0082) ***
Male	0.0358	0.0345	0.0348	0.0338	0.0253	0.0262	0.0235	0.0230	0.0233	0.0209	0.0183	0.0168
	(0.0077) ***	(0.0074) **	(0.0075) **	(0.0075) **	(0.0076) ***	(0.0073) **	(0.0073) *	(0.0074) *	(0.0075) **	(0.0073) *	(0.0075) .	(0.0075) .
Between 25 and 34	0.0237	0.0171	0.0149	0.0129	0.0576	0.0430	0.0474	0.0454	0.0879	0.0742	0.0845	0.0838
	(0.0126) .	(0.0121)	(0.0124)	(0.0124)	(0.0152) ***	(0.0146) *	(0.0147) *	(0.0148) *	(0.0173) ***	(0.0169) **	(0.0170) **	(0.0171) **
Between 35 and 44	0.0303	0.0215	0.0163	0.0141	0.0609	0.0435	0.0416	0.0393	0.0724	0.0588	0.0620	0.0605
	(0.0094) **	(0.0090) .	(0.0091)	(0.0091)	(0.0100) ***	(0.0096) **	(0.0097) **	(0.0097) **	(0.0106) ***	(0.0103) **	(0.0104) **	(0.0105) **
Between 45 and 54	0.0283	0.0164	0.0069	0.0031	0.0423	0.0230	0.0155	0.0130	0.0680	0.0503	0.0505	0.0468
	(0.0107) **	(0.0103)	(0.0106)	(0.0106)	(0.0118) ***	(0.0114) .	(0.0115)	(0.0115)	(0.0129) ***	(0.0125) **	(0.0125) **	(0.0126) **
Between 55 and 59	-0.0121	-0.0170	-0.0181	-0.0202	-0.0084	-0.0119	-0.0153	-0.0178	0.0074	-0.0029	-0.0033	-0.0050
	(0.0101)	(0.0097)	(0.0099)	(0.0099) .	(0.0103)	(0.0099)	(0.0100)	(0.0100)	(0.0110)	(0.0107)	(0.0107)	(0.0107)
Married filing Jointly	0.4535	0.3853	0.3588	0.3488	0.4200	0.3554	0.3428	0.3344	0.4433	0.3867	0.3872	0.3747
	(0.0211) ***	(0.0205) ***	(0.0215) ***	(0.0217) ***	(0.0203) ***	(0.0197) ***	(0.0207) ***	(0.0208) ***	(0.0203) ***	(0.0199) ***	(0.0208) ***	(0.0209) ***
Head of Household	-0.0291	-0.0049	-0.0149	-0.0114	-0.0756	-0.0479	-0.0508	-0.0476	-0.0432	-0.0190	-0.0249	-0.0219
	(0.0261)	(0.0251)	(0.0266)	(0.0267)	(0.0253) **	(0.0244) .	(0.0257) .	(0.0258)	(0.0255) .	(0.0248)	(0.0260)	(0.0263)
1 to 3 dependents	0.0267	0.0338	0.0291	0.0284	0.0500	0.0599	0.0599	0.0604	0.0161	0.0252	0.0220	0.0234
	(0.0117) *	(0.0112) *	(0.0115) .	(0.0115) .	(0.0114) ***	(0.0110) **	(0.0112) **	(0.0112) **	(0.0113)	(0.0110) .	(0.0112) .	(0.0112) .
4 or more dependents	0.0106	0.0243	-0.0024	-0.0051	0.0315	0.0498	0.0276	0.0263	-0.0002	0.0096	-0.0051	-0.0057
	(0.0184)	(0.0177)	(0.0183)	(0.0183)	(0.0180) .	(0.0173) *	(0.0178)	(0.0179)	(0.0179)	(0.0174)	(0.0179)	(0.0180)
Using a paid preparer	0.3935	0.0352	0.0776	0.0887	0.8816	-0.0054	0.0223	0.0335	-0.3788	-0.1385	-0.1713	-0.1644
	(0.2165) .	(0.0466)	(0.0543)	(0.0552)	(0.2309) ***	(0.0463)	(0.0541)	(0.0549)	(0.2143) .	(0.0458) *	(0.0526) *	(0.0538) *
Squared(Using a paid preparer)	-0.3745	-0.0023	0.0522	0.0513	-0.9186	0.0253	0.0786	0.0758	0.4185	0.1531	0.2334	0.2339
	(0.2321)	(0.0467)	(0.0530)	(0.0538)	(0.2475) ***	(0.0464)	(0.0525)	(0.0533)	(0.2302) .	(0.0458) **	(0.0513) **	(0.0523) **
ρ		0.3666		0.6594		0.3588		0.7114		0.3092		0.6286
		(0.0149) ***		(0.0327) ***		(0.0146) ***		(0.0294) ***		(0.0152) ***		(0.0348) ***
λ			0.8188				0.8202				0.7589	
			(0.0219) ***				(0.0218) ***				(0.0265) ***	

Footnotes at the end of table.

Table D.2: Subgroup 1 Model Estimates, 2002-2013, cont.

	2005				2006				2007			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Lag(Making Retirement Contributions)				-(0.2924)				-(0.1954)				-(0.0766)
				(0.1677)				(0.1536)				(0.1606)
Lag(Squared(Making Retirement Contributions))				(0.0010)				-(0.2068)				-(0.2375)
				(0.1668)				(0.1543)				(0.1595)
Lag(Total Adjusted Gross Income)				-(0.0820)				-(0.0100)				-(0.1422)
				(0.0839)				(0.0872)				(0.0882)
Lag(Squared(Total Adjusted Gross Income))				(0.0955)				(0.0250)				(0.1409)
				(0.0918)				(0.0953)				(0.0953)
Lag(Qualifying for 20 Percent Credit)				-(0.0253)				-(0.0399)				-(0.0350)
				(0.0536)				(0.0550)				(0.0494)
Lag(Qualifying for 10 Percent Credit)				-(0.0886)				-(0.0343)				-(0.1064)
				(0.0327) *				(0.0343)				(0.0368) *
Lag(90 Percent or more of income from SE[3])				(0.0600)				(0.0421)				(0.0962)
				(0.0253) .				(0.0248)				(0.0242) **
Lag(Male)				-(0.0463)				-(0.0151)				-(0.0323)
				(0.0403)				(0.0380)				(0.0394)
Lag(Between 25 and 34)				-(0.0580)				-(0.0461)				-(0.0741)
				(0.0660)				(0.0878)				(0.0951)
Lag(Between 35 and 44)				-(0.0076)				-(0.0234)				-(0.0360)
				(0.0521)				(0.0539)				(0.0575)
Lag(Between 45 and 54)				-(0.0427)				(0.0251)				-(0.0558)
				(0.0569)				(0.0721)				(0.0781)
Lag(Between 55 and 59)				-(0.0176)				-(0.0410)				(0.0305)
				(0.0583)				(0.0670)				(0.0684)
Lag(Married filing Jointly)				-(0.0976)				-(0.1229)				-(0.2163)
				(0.0952)				(0.0928)				(0.0957) .
Lag(Head of Household)				-(0.0082)				(0.0403)				-(0.0747)
				(0.1122)				(0.1088)				(0.1170)
Lag(1 to 3 dependents)				-(0.0349)				-(0.0829)				-(0.0058)
				(0.0529)				(0.0529)				(0.0537)
Lag(4 or more dependents)				(0.0451)				-(0.0133)				(0.0552)
				(0.0795)				(0.0811)				(0.0822)
Lag(Using a paid preparer)				(0.0269)				(0.0420)				(0.2173)
				(0.1338)				(0.1316)				(0.1319)
Lag(Squared(Using a paid preparer))				-(0.1668)				-(0.1477)				-(0.2961)
				(0.1337)				(0.1316)				(0.1317) .

Footnotes at the end of table.

Table D.2: Subgroup 1 Model Estimates, 2002-2013, cont.

	2005				2006				2007			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
N	7,455	7,455	7,455	7,455	7,453	7,453	7,453	7,453	7,476	7,476	7,476	7,476
AIC	13,153	12,598	12,481	12,438	12,969	12,423	12,245	12,228	12,778	12,390	12,269	12,243
Moran I Residuals	0.0872 ***	0.0254 ***	-0.0032	-0.0005	0.0961 ***	0.0340 ***	-0.0030	-0.0009	0.0759 ***	0.0283 ***	-0.0044	-0.0028
Log Likelihood		-6,277.85	-6,219.56	-6,179.78		-6,190.26	-6,101.42	-6,074.91		-6,173.90	-6,113.58	-6,082.50
Lagrange Multiplier												
LMerr	1,879 ***				2,280 ***				1,432 ***			
LMlag	887 ***				858 ***				565 ***			
RLMerr	1,295 ***				1,675 ***				1,046 ***			
RLMlag	304 ***				252 ***				178 ***			

Footnotes at the end of table.

Table D.2: Subgroup 1 Model Estimates, 2002-2013, cont.

	2008				2009				2010			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Intercept	-0.0890	0.0020	0.0142	0.0035	-0.0624	0.0019	0.0109	0.0039	0.0010	0.0013	0.0087	0.0032
	(0.0291) **	(0.0065)	(0.0285)	(0.0065)	(0.0297) *	(0.0062)	(0.0223)	(0.0061)	(0.0284)	(0.0062)	(0.0224)	(0.0062)
Making Retirement Contributions	1.0306	0.9144	0.9541	0.9421	0.8909	0.8052	0.8406	0.8307	0.6718	0.6000	0.6197	0.6116
	(0.0287) ***	(0.0285) ***	(0.0284) ***	(0.0285) ***	(0.0274) ***	(0.0272) ***	(0.0274) ***	(0.0275) ***	(0.0275) ***	(0.0270) ***	(0.0273) ***	(0.0274) ***
Squared(Making Retirement Contributions)	-0.3266	-0.2754	-0.2939	-0.2868	-0.1524	-0.1224	-0.1371	-0.1327	0.0666	0.0889	0.0916	0.0962
	(0.0282) ***	(0.0275) ***	(0.0275) ***	(0.0275) ***	(0.0269) ***	(0.0263) **	(0.0265) **	(0.0265) **	(0.0271) *	(0.0263) **	(0.0266) **	(0.0267) **
Total Adjusted Gross Income	-0.0561	-0.0610	-0.0489	-0.0497	-0.0342	-0.0367	-0.0195	-0.0183	-0.0329	-0.0329	-0.0155	-0.0134
	(0.0192) **	(0.0186) *	(0.0195) .	(0.0195) .	(0.0179) .	(0.0174) .	(0.0182)	(0.0184)	(0.0176) .	(0.0172)	(0.0179)	(0.0180)
Squared(Total Adjusted Gross Income)	0.0271	0.0280	0.0190	0.0199	0.0116	0.0127	0.0030	0.0032	0.0037	0.0050	-0.0031	-0.0029
	(0.0180)	(0.0175)	(0.0177)	(0.0178)	(0.0168)	(0.0164)	(0.0166)	(0.0166)	(0.0166)	(0.0162)	(0.0164)	(0.0164)
Qualifying for 20 Percent Credit	-0.0024	-0.0048	-0.0023	-0.0011	-0.0113	-0.0108	-0.0094	-0.0101	0.0039	0.0047	0.0050	0.0065
	(0.0073)	(0.0071)	(0.0070)	(0.0070)	(0.0072)	(0.0070)	(0.0070)	(0.0070)	(0.0068)	(0.0066)	(0.0066)	(0.0066)
Qualifying for 10 Percent Credit	-0.0169	-0.0009	0.0087	0.0149	-0.0099	0.0011	0.0044	0.0098	0.0034	0.0124	0.0178	0.0226
	(0.0086) .	(0.0084)	(0.0091)	(0.0093)	(0.0082)	(0.0080)	(0.0087)	(0.0089)	(0.0081)	(0.0079)	(0.0085) .	(0.0087) *
90 percent or more of income from SE [3]	-0.0527	-0.0307	-0.0360	-0.0314	-0.0623	-0.0406	-0.0519	-0.0475	-0.0879	-0.0672	-0.0798	-0.0754
	(0.0078) ***	(0.0076) **	(0.0082) **	(0.0082) **	(0.0074) ***	(0.0073) **	(0.0078) ***	(0.0079) **	(0.0073) ***	(0.0073) ***	(0.0078) ***	(0.0079) ***
Male	0.0340	0.0300	0.0248	0.0232	0.0303	0.0310	0.0323	0.0321	0.0189	0.0192	0.0209	0.0201
	(0.0078) ***	(0.0075) **	(0.0077) *	(0.0077) *	(0.0073) ***	(0.0071) **	(0.0072) **	(0.0073) **	(0.0073) **	(0.0071) *	(0.0073) *	(0.0073) *
Between 25 and 34	0.0867	0.0828	0.0990	0.1015	0.1062	0.0993	0.1034	0.1029	0.1039	0.0951	0.1007	0.1007
	(0.0190) ***	(0.0184) **	(0.0188) **	(0.0189) **	(0.0148) ***	(0.0145) ***	(0.0149) ***	(0.0150) ***	(0.0131) ***	(0.0128) ***	(0.0132) ***	(0.0133) ***
Between 35 and 44	0.0930	0.0779	0.0807	0.0805	0.0859	0.0760	0.0793	0.0781	0.1087	0.0982	0.1028	0.1021
	(0.0113) ***	(0.0110) ***	(0.0112) ***	(0.0112) ***	(0.0096) ***	(0.0094) ***	(0.0096) ***	(0.0096) ***	(0.0091) ***	(0.0089) ***	(0.0091) ***	(0.0092) ***
Between 45 and 54	0.0534	0.0374	0.0369	0.0346	0.1029	0.0889	0.0892	0.0875	0.1115	0.0966	0.1020	0.0994
	(0.0137) ***	(0.0133) *	(0.0134) *	(0.0134) *	(0.0116) ***	(0.0113) ***	(0.0114) ***	(0.0114) ***	(0.0105) ***	(0.0103) ***	(0.0104) ***	(0.0105) ***
Between 55 and 59	0.0147	0.0110	0.0098	0.0095	0.0246	0.0186	0.0213	0.0188	0.0712	0.0621	0.0629	0.0610
	(0.0118)	(0.0115)	(0.0114)	(0.0114)	(0.0103) *	(0.0100)	(0.0100) .	(0.0100)	(0.0099) ***	(0.0097) ***	(0.0097) **	(0.0097) ***
Married filing Jointly	0.4123	0.3589	0.3429	0.3323	0.3748	0.3365	0.3258	0.3172	0.3508	0.3010	0.2837	0.2728
	(0.0210) ***	(0.0205) ***	(0.0211) ***	(0.0211) ***	(0.0190) ***	(0.0187) ***	(0.0192) ***	(0.0193) ***	(0.0182) ***	(0.0180) ***	(0.0186) ***	(0.0187) ***
Head of Household	-0.0455	-0.0180	-0.0089	-0.0018	-0.0649	-0.0431	-0.0511	-0.0438	-0.1078	-0.0996	-0.1133	-0.1106
	(0.0257) .	(0.0249)	(0.0260)	(0.0261)	(0.0222) **	(0.0217) .	(0.0225) .	(0.0227)	(0.0216) ***	(0.0211) **	(0.0220) **	(0.0222) **
1 to 3 dependents	0.0385	0.0486	0.0472	0.0471	0.0282	0.0366	0.0351	0.0356	0.0309	0.0437	0.0433	0.0445
	(0.0116) ***	(0.0112) **	(0.0115) **	(0.0115) **	(0.0109) **	(0.0106) **	(0.0109) *	(0.0109) *	(0.0108) **	(0.0106) **	(0.0108) **	(0.0109) **
4 or more dependents	0.0207	0.0344	0.0092	0.0066	0.0037	0.0187	0.0067	0.0062	0.0223	0.0399	0.0244	0.0232
	(0.0180)	(0.0174) .	(0.0182)	(0.0183)	(0.0156)	(0.0152)	(0.0158)	(0.0160)	(0.0155)	(0.0152) *	(0.0159)	(0.0160)
Using a paid preparer	0.6099	0.0373	0.0793	0.0960	0.3636	-0.0516	-0.0840	-0.0809	-0.1033	-0.0652	-0.0492	-0.0403
	(0.2172) **	(0.0462)	(0.0538)	(0.0549)	(0.2263)	(0.0446)	(0.0520)	(0.0537)	(0.2137)	(0.0436)	(0.0505)	(0.0521)
Squared(Using a paid preparer)	-0.6244	-0.0132	0.0028	-0.0072	-0.3541	0.0815	0.1404	0.1373	0.1304	0.0765	0.0880	0.0818
	(0.2334) **	(0.0461)	(0.0523)	(0.0532)	(0.2433)	(0.0445)	(0.0500) *	(0.0512) *	(0.2296)	(0.0434)	(0.0485)	(0.0496)
ρ		0.3247		0.5904		0.2685		0.5890		0.2641		0.6038
		(0.0151) ***		(0.0370) ***		(0.0145) ***		(0.0372) ***		(0.0148) ***		(0.0364) ***
λ			0.7730				0.7254				0.7251	
			(0.0255) ***				(0.0290) ***				(0.0290) ***	

Footnotes at the end of table.

Table D.2: Subgroup 1 Model Estimates, 2002-2013, cont.

	2008				2009				2010			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Lag(Making Retirement Contributions)				-(0.4899)				-(0.4062)				-(0.2189)
				(0.1801) *				(0.1542) *				(0.1429)
Lag(Squared(Making Retirement Contributions))				(0.2036)				(0.0734)				-(0.1343)
				(0.1793)				(0.1543)				(0.1420)
Lag(Total Adjusted Gross Income)				(0.0022)				-(0.0158)				-(0.0793)
				(0.0886)				(0.0816)				(0.0833)
Lag(Squared(Total Adjusted Gross Income))				(0.0156)				(0.0162)				(0.0732)
				(0.0948)				(0.0868)				(0.0879)
Lag(Qualifying for 20 Percent Credit)				(0.0541)				-(0.0502)				(0.0434)
				(0.0600)				(0.0600)				(0.0538)
Lag(Qualifying for 10 Percent Credit)				(0.0028)				-(0.0517)				-(0.0497)
				(0.0372)				(0.0358)				(0.0367)
Lag(90 Percent or more of income from SE[3])				(0.0421)				(0.0663)				(0.0973)
				(0.0269)				(0.0263) .				(0.0265) **
Lag(Male)				(0.0074)				-(0.0362)				-(0.0293)
				(0.0401)				(0.0390)				(0.0370)
Lag(Between 25 and 34)				-(0.1058)				-(0.1268)				-(0.1218)
				(0.1079)				(0.0848)				(0.0718)
Lag(Between 35 and 44)				-(0.0059)				-(0.0823)				-(0.0519)
				(0.0617)				(0.0527)				(0.0421)
Lag(Between 45 and 54)				-(0.0164)				-(0.0249)				-(0.1154)
				(0.0834)				(0.0727)				(0.0639)
Lag(Between 55 and 59)				(0.0138)				-(0.1402)				-(0.0790)
				(0.0868)				(0.0741)				(0.0661)
Lag(Married filing Jointly)				-(0.1650)				-(0.0517)				-(0.0360)
				(0.1005)				(0.0870)				(0.0792)
Lag(Head of Household)				-(0.1198)				(0.0713)				(0.0975)
				(0.1207)				(0.0995)				(0.0940)
Lag(1 to 3 dependents)				-(0.0255)				-(0.0441)				-(0.0648)
				(0.0568)				(0.0536)				(0.0480)
Lag(4 or more dependents)				(0.0903)				(0.0017)				-(0.0008)
				(0.0856)				(0.0721)				(0.0691)
Lag(Using a paid preparer)				-(0.0634)				(0.1374)				(0.0625)
				(0.1312)				(0.1274)				(0.1283)
Lag(Squared(Using a paid preparer))				-(0.0201)				-(0.1770)				-(0.1004)
				(0.1308)				(0.1277)				(0.1282)

Footnotes at the end of table.

Table D.2: Subgroup 1 Model Estimates, 2002-2013, cont.

	2008				2009				2010			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
N	7,449	7,449	7,449	7,449	7,438	7,438	7,438	7,438	7,424	7,424	7,424	7,424
AIC	13,083	12,656	12,587	12,541	12,146	11,824	11,785	11,732	12,180	11,885	11,773	11,755
Moran I Residuals	0.0695 ***	0.0217 ***	-0.0039	-0.0010	0.0614 ***	0.0235 ***	-0.0023	-0.0007	0.0647 ***	0.0297 ***	-0.0022	-0.0006
Log Likelihood		-6,306.81	-6,272.43	-6,231.36		-5,890.90	-5,858.16	-5,827.24		-5,921.36	-5,865.50	-5,838.41
Lagrange Multiplier												
LMerr	1,194 ***				929 ***				1,026 ***			
LMlag	639 ***				453 ***				400 ***			
RLMerr	806 ***				662 ***				763 ***			
RLMlag	251 ***				185 ***				137 ***			

Footnotes at the end of table.

Table D.2: Subgroup 1 Model Estimates, 2002-2013, cont.

	2011				2012				2013			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Intercept	-0.0984	0.0006	0.0061	0.0026	-0.0161	0.0005	0.0053	0.0022	-0.1087	0.0000	0.0033	0.0019
	(0.0265) ***	(0.0063)	(0.0186)	(0.0062)	(0.0298)	(0.0064)	(0.0167)	(0.0064)	(0.0287) ***	(0.0066)	(0.0181)	(0.0066)
Making Retirement Contributions	0.7768	0.7209	0.7444	0.7337	0.8386	0.7775	0.8082	0.7974	0.8750	0.8123	0.8458	0.8361
	(0.0287) ***	(0.0287) ***	(0.0292) ***	(0.0294) ***	(0.0291) ***	(0.0294) ***	(0.0296) ***	(0.0299) ***	(0.0295) ***	(0.0298) ***	(0.0299) ***	(0.0301) ***
Squared(Making Retirement Contributions)	-0.0265	-0.0096	-0.0141	-0.0065	-0.0690	-0.0417	-0.0526	-0.0456	-0.1076	-0.0830	-0.0954	-0.0906
	(0.0284)	(0.0281)	(0.0286)	(0.0287)	(0.0287) *	(0.0284)	(0.0288)	(0.0289)	(0.0291) ***	(0.0289) *	(0.0291) *	(0.0292) *
Total Adjusted Gross Income	-0.0551	-0.0568	-0.0411	-0.0397	-0.0462	-0.0504	-0.0309	-0.0254	-0.0264	-0.0338	-0.0168	-0.0145
	(0.0173) **	(0.0171) **	(0.0177) .	(0.0178) .	(0.0180) *	(0.0178) *	(0.0184)	(0.0186)	(0.0189)	(0.0187)	(0.0193)	(0.0195)
Squared(Total Adjusted Gross Income)	0.0168	0.0201	0.0123	0.0132	0.0076	0.0123	0.0002	-0.0016	0.0032	0.0105	0.0002	0.0007
	(0.0162)	(0.0160)	(0.0161)	(0.0161)	(0.0167)	(0.0166)	(0.0167)	(0.0168)	(0.0177)	(0.0175)	(0.0176)	(0.0177)
Qualifying for 20 Percent Credit	0.0061	0.0054	0.0087	0.0088	-0.0051	-0.0069	-0.0072	-0.0050	-0.0239	-0.0263	-0.0238	-0.0223
	(0.0068)	(0.0067)	(0.0067)	(0.0067)	(0.0071)	(0.0070)	(0.0070)	(0.0070)	(0.0073) **	(0.0072) **	(0.0071) **	(0.0071) *
Qualifying for 10 Percent Credit	0.0225	0.0299	0.0416	0.0474	0.0058	0.0091	0.0117	0.0189	-0.0066	-0.0052	0.0001	0.0060
	(0.0082) **	(0.0081) **	(0.0086) **	(0.0088) **	(0.0084)	(0.0083)	(0.0089)	(0.0091) .	(0.0086)	(0.0085)	(0.0090)	(0.0092)
90 percent or more of income from SE [3]	-0.0720	-0.0549	-0.0593	-0.0547	-0.0652	-0.0555	-0.0579	-0.0523	-0.0517	-0.0393	-0.0335	-0.0264
	(0.0074) ***	(0.0074) ***	(0.0079) ***	(0.0081) ***	(0.0075) ***	(0.0075) ***	(0.0082) ***	(0.0084) **	(0.0077) ***	(0.0077) **	(0.0084) **	(0.0087) *
Male	0.0137	0.0113	0.0117	0.0101	-0.0092	-0.0092	-0.0123	-0.0140	0.0159	0.0124	0.0089	0.0075
	(0.0073) .	(0.0072)	(0.0073)	(0.0074)	(0.0075)	(0.0074)	(0.0075)	(0.0075)	(0.0077) *	(0.0076)	(0.0077)	(0.0078)
Between 25 and 34	0.0947	0.0835	0.0946	0.0963	0.0836	0.0803	0.0885	0.0909	0.0931	0.0919	0.1011	0.1042
	(0.0116) ***	(0.0115) ***	(0.0118) ***	(0.0119) ***	(0.0108) ***	(0.0107) ***	(0.0110) ***	(0.0111) ***	(0.0101) ***	(0.0100) ***	(0.0103) ***	(0.0104) ***
Between 35 and 44	0.1115	0.1028	0.1037	0.1023	0.1104	0.1045	0.1053	0.1045	0.0902	0.0837	0.0836	0.0822
	(0.0088) ***	(0.0087) ***	(0.0089) ***	(0.0090) ***	(0.0087) ***	(0.0086) ***	(0.0088) ***	(0.0089) ***	(0.0089) ***	(0.0088) ***	(0.0092) ***	(0.0093) ***
Between 45 and 54	0.1179	0.1023	0.1075	0.1049	0.0997	0.0894	0.0890	0.0880	0.1023	0.0924	0.0916	0.0905
	(0.0097) ***	(0.0096) ***	(0.0097) ***	(0.0098) ***	(0.0093) ***	(0.0093) ***	(0.0094) ***	(0.0095) ***	(0.0090) ***	(0.0089) ***	(0.0091) ***	(0.0091) ***
Between 55 and 59	0.0573	0.0500	0.0519	0.0492	0.0524	0.0487	0.0481	0.0466	0.0615	0.0580	0.0569	0.0559
	(0.0095) ***	(0.0094) **	(0.0094) **	(0.0094) **	(0.0092) ***	(0.0091) **	(0.0091) **	(0.0091) **	(0.0091) ***	(0.0090) ***	(0.0090) **	(0.0090) ***
Married filing Jointly	0.3404	0.2976	0.2831	0.2692	0.3220	0.2963	0.2910	0.2796	0.3464	0.3213	0.3128	0.3052
	(0.0168) ***	(0.0168) ***	(0.0173) ***	(0.0175) ***	(0.0182) ***	(0.0182) ***	(0.0185) ***	(0.0186) ***	(0.0181) ***	(0.0180) ***	(0.0182) ***	(0.0183) ***
Head of Household	-0.1077	-0.1143	-0.1183	-0.1168	-0.1023	-0.0981	-0.0913	-0.0850	-0.0524	-0.0512	-0.0471	-0.0398
	(0.0195) ***	(0.0192) ***	(0.0201) **	(0.0203) **	(0.0214) ***	(0.0212) **	(0.0220) **	(0.0223) **	(0.0211) *	(0.0209) .	(0.0217) .	(0.0220)
1 to 3 dependents	0.0190	0.0336	0.0315	0.0318	0.0464	0.0513	0.0518	0.0509	0.0384	0.0445	0.0503	0.0498
	(0.0102) .	(0.0101) **	(0.0103) *	(0.0104) *	(0.0105) ***	(0.0104) **	(0.0106) **	(0.0106) **	(0.0108) ***	(0.0107) **	(0.0109) **	(0.0110) **
4 or more dependents	0.0357	0.0561	0.0416	0.0392	0.0412	0.0495	0.0331	0.0281	0.0449	0.0552	0.0469	0.0437
	(0.0144) *	(0.0143) **	(0.0149) *	(0.0151) *	(0.0151) **	(0.0150) **	(0.0156) .	(0.0159)	(0.0154) **	(0.0153) **	(0.0160) *	(0.0163) *
Using a paid preparer	0.6610	-0.0621	0.0071	0.0346	-0.0139	-0.1105	-0.1095	-0.0959	0.6275	-0.1610	-0.1635	-0.1543
	(0.1994) ***	(0.0429)	(0.0491)	(0.0512)	(0.2292)	(0.0447) .	(0.0508) .	(0.0534) .	(0.2165) **	(0.0453) **	(0.0514) *	(0.0536) *
Squared(Using a paid preparer)	-0.6780	0.0918	0.0586	0.0418	0.0312	0.1284	0.1503	0.1440	-0.6418	0.1935	0.2149	0.2070
	(0.2143) **	(0.0427) .	(0.0472)	(0.0486)	(0.2462)	(0.0446) *	(0.0493) *	(0.0512) *	(0.2318) **	(0.0452) **	(0.0500) **	(0.0517) **
ρ		0.2093		0.5477		0.1844		0.5029		0.2135		0.5332
		(0.0157) ***		(0.0400) ***		(0.0166) ***		(0.0424) ***		(0.0174) ***		(0.0407) ***
λ			0.6664				0.6198				0.6386	
			(0.0330) ***				(0.0360) ***				(0.0348) ***	

Footnotes at the end of table.

Table D.2: Subgroup 1 Model Estimates, 2002-2013, cont.

	2011				2012				2013			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Lag(Making Retirement Contributions)				-(0.3219)				-(0.4351)				-(0.3597)
				(0.1436) .				(0.1538) *				(0.1665) .
Lag(Squared(Making Retirement Contributions))				(0.0143)				(0.0799)				-(0.0319)
				(0.1414)				(0.1526)				(0.1645)
Lag(Total Adjusted Gross Income)				-(0.0911)				(0.0191)				-(0.0106)
				(0.0840)				(0.0880)				(0.0933)
Lag(Squared(Total Adjusted Gross Income))				(0.0713)				-(0.0095)				(0.0352)
				(0.0880)				(0.0921)				(0.0965)
Lag(Qualifying for 20 Percent Credit)				-(0.0689)				(0.1210)				(0.0519)
				(0.0546)				(0.0559) .				(0.0604)
Lag(Qualifying for 10 Percent Credit)				-(0.1226)				(0.0386)				(0.0501)
				(0.0372) **				(0.0400)				(0.0411)
Lag(90 Percent or more of income from SE[3])				(0.0575)				(0.0177)				-(0.0467)
				(0.0283) .				(0.0274)				(0.0283)
Lag(Male)				-(0.0463)				(0.0016)				(0.0232)
				(0.0397)				(0.0411)				(0.0420)
Lag(Between 25 and 34)				-(0.1178)				-(0.0432)				-(0.0449)
				(0.0640)				(0.0604)				(0.0585)
Lag(Between 35 and 44)				-(0.0277)				-(0.0099)				-(0.0174)
				(0.0418)				(0.0429)				(0.0420)
Lag(Between 45 and 54)				-(0.0815)				(0.1044)				(0.1029)
				(0.0570)				(0.0556)				(0.0527)
Lag(Between 55 and 59)				-(0.1203)				-(0.0036)				-(0.0089)
				(0.0647)				(0.0686)				(0.0693)
Lag(Married filing Jointly)				-(0.0304)				-(0.0532)				(0.0242)
				(0.0720)				(0.0831)				(0.0872)
Lag(Head of Household)				(0.0900)				(0.0071)				(0.1446)
				(0.0856)				(0.0957)				(0.0938)
Lag(1 to 3 dependents)				-(0.0874)				-(0.0802)				-(0.1618)
				(0.0493)				(0.0517)				(0.0536) *
Lag(4 or more dependents)				-(0.0516)				(0.0130)				-(0.1098)
				(0.0662)				(0.0679)				(0.0664)
Lag(Using a paid preparer)				-(0.0602)				(0.1863)				(0.2008)
				(0.1245)				(0.1287)				(0.1281)
Lag(Squared(Using a paid preparer))				-(0.0051)				-(0.2214)				-(0.2264)
				(0.1251)				(0.1289)				(0.1282)

Footnotes at the end of table.

Table D.2: Subgroup 1 Model Estimates, 2002-2013, cont.

	2011				2012				2013			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
N	7,397	7,397	7,397	7,397	7,373	7,373	7,373	7,373	7,355	7,355	7,355	7,355
AIC	12,117	11,950	11,846	11,830	12,324	12,207	12,107	12,090	12,744	12,604	12,497	12,484
Moran I Residuals	0.0486 ***	0.0256 ***	-0.0025	-0.0008	0.0430 ***	0.0253 ***	-0.0021	-0.0012	0.0476 ***	0.0264 ***	-0.0023	-0.0011
Log Likelihood		-5,953.83	-5,902.20	-5,876.04		-6,082.54	-6,032.37	-6,006.07		-6,280.81	-6,227.69	-6,202.86
Lagrange Multiplier												
LMerr	574 ***				448 ***				545 ***			
LMlag	218 ***				146 ***				178 ***			
RLMerr	428 ***				342 ***				410 ***			
RLMlag	72 ***				40 ***				43 ***			

Standard Errors are in paretheses below parameter estimates
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Standard Errors in Paretheses below parameter estimates
[1] All parameters reflect the percent for a given ZIP Code except Total AGI, which is the sum of reported AGI
[2] Lagged parameters are represented by θ in the SDM spatial model.
[3] SE - Self-Employed
Source: author's computation

Table D.3: Subgroup 2 Model Estimates, 2002-2013

	2002				2003				2004			
Parameter [1][2]	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Intercept	0.0000	0.0057	0.0310	0.0062	0.0000	0.0045	0.0181	0.0030	0.0000	0.0049	0.0185	0.0046
	(0.0109)	(0.0102)	(0.0611)	(0.0101)	(0.0106)	(0.0101)	(0.0540)	(0.0100)	(0.0105)	(0.0100)	(0.0450)	(0.0100)
Total Adjusted Gross Income	0.1031	0.0612	0.0468	0.0453	0.1446	0.1126	0.0826	0.0806	0.1474	0.1221	0.0939	0.0954
	(0.0153) ***	(0.0144) **	(0.0154) *	(0.0154) *	(0.0277) ***	(0.0264) **	(0.0277) *	(0.0279) *	(0.0268) ***	(0.0256) **	(0.0270) **	(0.0272) **
Squared(Total Adjusted Gross Income)	-0.0713	-0.0498	-0.0404	-0.0417	-0.0844	-0.0717	-0.0512	-0.0505	-0.0983	-0.0883	-0.0712	-0.0728
	(0.0142) ***	(0.0133) **	(0.0135) *	(0.0136) *	(0.0265) **	(0.0252) *	(0.0257) .	(0.0257) .	(0.0256) ***	(0.0245) **	(0.0250) **	(0.0251) *
Qualifying for 20 Percent Credit	0.0185	0.0134	0.0125	0.0108	0.0052	0.0022	0.0011	0.0019	0.0209	0.0207	0.0186	0.0187
	(0.0120)	(0.0112)	(0.0112)	(0.0112)	(0.0113)	(0.0108)	(0.0106)	(0.0107)	(0.0116) .	(0.0111)	(0.0110)	(0.0111)
Qualifying for 10 Percent Credit	0.0911	0.1057	0.0968	0.0974	0.0145	0.0342	0.0245	0.0262	-0.0040	0.0226	0.0104	0.0121
	(0.0132) ***	(0.0124) ***	(0.0134) ***	(0.0135) ***	(0.0125)	(0.0120) *	(0.0126)	(0.0126) .	(0.0125)	(0.0121)	(0.0125)	(0.0126)
90 percent or more of income from SE [3]	-0.0093	0.0106	-0.0056	-0.0043	-0.0088	0.0292	0.0100	0.0116	-0.0122	0.0180	0.0068	0.0098
	(0.0116)	(0.0110)	(0.0112)	(0.0113)	(0.0113)	(0.0109) *	(0.0111)	(0.0111)	(0.0112)	(0.0108)	(0.0109)	(0.0109)
Male	0.0582	0.0597	0.0800	0.0811	0.0912	0.0812	0.0985	0.0983	0.0406	0.0268	0.0395	0.0382
	(0.0127) ***	(0.0119) **	(0.0122) ***	(0.0122) ***	(0.0120) ***	(0.0115) ***	(0.0118) ***	(0.0118) ***	(0.0119) ***	(0.0115) .	(0.0117) **	(0.0117) *
Between 25 and 34	0.0318	0.0137	0.0195	0.0153	0.0970	0.0798	0.0874	0.0831	0.0283	-0.0028	0.0129	0.0054
	(0.0178) .	(0.0167)	(0.0169)	(0.0169)	(0.0180) ***	(0.0172) **	(0.0172) **	(0.0173) **	(0.0182)	(0.0175)	(0.0176)	(0.0176)
Between 35 and 44	0.0104	-0.0127	-0.0164	-0.0217	0.0414	0.0190	0.0195	0.0156	-0.0128	-0.0444	-0.0421	-0.0484
	(0.0168)	(0.0158)	(0.0161)	(0.0162)	(0.0169) *	(0.0162)	(0.0162)	(0.0163)	(0.0168)	(0.0161) *	(0.0162) *	(0.0162) *
Between 45 and 54	-0.0493	-0.0731	-0.0768	-0.0814	0.0171	-0.0094	-0.0028	-0.0056	-0.0028	-0.0358	-0.0320	-0.0381
	(0.0173) **	(0.0162) **	(0.0168) **	(0.0169) **	(0.0168)	(0.0160)	(0.0163)	(0.0164)	(0.0169)	(0.0162) .	(0.0164)	(0.0165) .
Between 55 and 59	0.0155	0.0029	-0.0008	-0.0060	0.0287	0.0206	0.0182	0.0145	-0.0193	-0.0369	-0.0342	-0.0401
	(0.0145)	(0.0136)	(0.0138)	(0.0139)	(0.0150) .	(0.0142)	(0.0143)	(0.0144)	(0.0150)	(0.0143) *	(0.0144) .	(0.0145) *
Married filing Jointly	0.3862	0.3177	0.3249	0.3285	0.3214	0.2501	0.2384	0.2385	0.4496	0.3564	0.3599	0.3548
	(0.0256) ***	(0.0241) ***	(0.0249) ***	(0.0250) ***	(0.0248) ***	(0.0237) ***	(0.0243) ***	(0.0244) ***	(0.0234) ***	(0.0226) ***	(0.0234) ***	(0.0235) ***
Head of Household	0.1931	0.2407	0.2474	0.2644	0.0646	0.1190	0.1037	0.1152	0.1976	0.2217	0.2238	0.2338
	(0.0307) ***	(0.0288) ***	(0.0301) ***	(0.0303) ***	(0.0299) *	(0.0286) **	(0.0293) **	(0.0295) **	(0.0287) ***	(0.0274) ***	(0.0284) ***	(0.0285) ***
1 to 3 dependents	-0.0215	-0.0060	0.0058	0.0056	0.0564	0.0658	0.0757	0.0750	0.1045	0.1027	0.1143	0.1170
	(0.0154)	(0.0144)	(0.0146)	(0.0146)	(0.0153) ***	(0.0146) **	(0.0146) **	(0.0147) **	(0.0150) ***	(0.0143) ***	(0.0145) ***	(0.0145) ***
4 or more dependents	-0.0384	0.0028	0.0190	0.0196	0.0306	0.0552	0.0561	0.0559	0.0260	0.0453	0.0562	0.0605
	(0.0205) .	(0.0193)	(0.0199)	(0.0200)	(0.0204)	(0.0195) *	(0.0200) *	(0.0201) *	(0.0201)	(0.0192) .	(0.0199) *	(0.0200) *
Using a paid preparer	0.2191	0.1610	0.1815	0.1740	0.4455	0.3770	0.4200	0.4134	0.3380	0.3396	0.3632	0.3579
	(0.0507) ***	(0.0476) **	(0.0502) **	(0.0503) **	(0.0510) ***	(0.0487) ***	(0.0515) ***	(0.0517) ***	(0.0504) ***	(0.0481) ***	(0.0506) ***	(0.0508) ***
Squared(Using a paid preparer)	-0.1002	-0.0643	-0.0074	0.0033	-0.2684	-0.2292	-0.1673	-0.1529	-0.1749	-0.2150	-0.1654	-0.1582
	(0.0504) *	(0.0474)	(0.0494)	(0.0495)	(0.0510) ***	(0.0486) **	(0.0507) **	(0.0508) *	(0.0504) ***	(0.0483) **	(0.0501) **	(0.0502) *
ρ		0.7627		0.7441		0.6886		0.7158		0.6454		0.6660
		(0.0202) ***		(0.0270) ***		(0.0222) ***		(0.0291) ***		(0.0238) ***		(0.0324) ***
λ			0.8345				0.8157				0.7780	
			(0.0206) ***				(0.0222) ***				(0.0251) ***	

Footnotes at the end of table.

Table D.3: Subgroup 2 Model Estimates, 2002-2013, cont.

	2002				2003				2004			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Lag(Total Adjusted Gross Income)				(0.0889)				(0.1436)				(0.2322)
				(0.0592)				(0.1216)				(0.1180) .
Lag(Squared(Total Adjusted Gross Income))				-(0.1354)				-(0.1283)				-(0.1580)
				(0.0876)				(0.1305)				(0.1243)
Lag(Qualifying for 20 Percent Credit)				-(0.0872)				(0.0721)				-(0.0348)
				(0.0912)				(0.0833)				(0.0766)
Lag(Qualifying for 10 Percent Credit)				-(0.1299)				-(0.0011)				-(0.0402)
				(0.0564) .				(0.0561)				(0.0561)
Lag(90 Percent or more of income from SE[3])				(0.0382)				-(0.0445)				(0.0106)
				(0.0617)				(0.0618)				(0.0628)
Lag(Male)				-(0.1553)				-(0.1703)				-(0.0504)
				(0.0553) *				(0.0548) *				(0.0540)
Lag(Between 25 and 34)				-(0.1840)				-(0.2422)				-(0.3265)
				(0.1077)				(0.1176) .				(0.1181) *
Lag(Between 35 and 44)				-(0.1534)				-(0.0610)				-(0.0843)
				(0.0972)				(0.1098)				(0.1089)
Lag(Between 45 and 54)				-(0.0417)				-(0.0984)				-(0.0675)
				(0.0903)				(0.1003)				(0.0960)
Lag(Between 55 and 59)				-(0.2207)				-(0.2267)				-(0.2025)
				(0.0957) .				(0.1012) .				(0.1052)
Lag(Married filing Jointly)				(0.3144)				(0.2301)				(0.0590)
				(0.1389) .				(0.1431)				(0.1259)
Lag(Head of Household)				(0.3031)				(0.1509)				-(0.0329)
				(0.1570)				(0.1721)				(0.1578)
Lag(1 to 3 dependents)				-(0.1801)				-(0.2117)				-(0.0253)
				(0.0963)				(0.1005) .				(0.0968)
Lag(4 or more dependents)				-(0.3239)				-(0.1561)				-(0.0254)
				(0.1146) *				(0.1234)				(0.1229)
Lag(Using a paid preparer)				(0.0194)				-(0.0841)				-(0.2319)
				(0.1843)				(0.1908)				(0.2009)
Lag(Squared(Using a paid preparer))				-(0.1904)				-(0.1494)				(0.0806)
				(0.1846)				(0.1908)				(0.2019)

Footnotes at the end of table.

Table D.3: Subgroup 2 Model Estimates, 2002-2013, cont.

	2002				2003				2004			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
N	7,469	7,469	7,469	7,469	7,468	7,468	7,468	7,468	7,462	7,462	7,462	7,462
AIC	20,249	19,420	19,347	19,321	19,847	19,221	19,109	19,090	19,751	19,161	19,112	19,092
Moran I Residuals	0.1171 ***	0.0057 *	-0.0034	-0.0015	0.0981 ***	0.0112 ***	-0.0030	-0.0012	0.0920 ***	0.0052 *	-0.0057	-0.0046
Log Likelihood		-9,691.12	-9,654.54	-9,625.38		-9,591.61	-9,535.29	-9,510.01		-9,561.56	-9,536.95	-9,510.92
Lagrange Multiplier												
Error	3,400 ***				2,387 ***				2,095 ***			
LMlag	2,612 ***				1,547 ***				1,493 ***			
RLMerr	866 ***				890 ***				710 ***			
RLMlag	78 ***				50 ***				108 ***			

Footnotes at the end of table.

Table D.3: Subgroup 2 Model Estimates, 2002-2013, cont.

	2005				2006				2007			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Intercept	0.0000	0.0058	0.0197	0.0051	0.0000	0.0048	0.0193	0.0055	0.0000	0.0041	0.0157	0.0054
	(0.0105)	(0.0101)	(0.0420)	(0.0101)	(0.0105)	(0.0101)	(0.0406)	(0.0101)	(0.0107)	(0.0104)	(0.0371)	(0.0103)
Total Adjusted Gross Income	0.1276	0.1268	0.1174	0.1186	0.0808	0.0787	0.0641	0.0642	0.0422	0.0434	0.0271	0.0265
	(0.0270) ***	(0.0260) **	(0.0275) **	(0.0276) **	(0.0268) **	(0.0258) *	(0.0273) .	(0.0274) .	(0.0238) .	(0.0230)	(0.0241)	(0.0242)
Squared(Total Adjusted Gross Income)	-0.0764	-0.0864	-0.0817	-0.0821	-0.0442	-0.0554	-0.0455	-0.0457	-0.0166	-0.0304	-0.0210	-0.0216
	(0.0258) **	(0.0248) **	(0.0255) *	(0.0255) *	(0.0257) .	(0.0247) .	(0.0253)	(0.0253)	(0.0228)	(0.0221)	(0.0224)	(0.0224)
Qualifying for 20 Percent Credit	0.0049	0.0101	0.0057	0.0061	0.0205	0.0245	0.0206	0.0195	0.0374	0.0383	0.0381	0.0340
	(0.0116)	(0.0111)	(0.0111)	(0.0111)	(0.0118) .	(0.0113) .	(0.0113)	(0.0113)	(0.0121) **	(0.0117) *	(0.0117) *	(0.0118) *
Qualifying for 10 Percent Credit	0.0029	0.0366	0.0230	0.0262	0.0100	0.0268	0.0119	0.0142	0.0514	0.0659	0.0542	0.0554
	(0.0125)	(0.0121) *	(0.0127)	(0.0127) .	(0.0125)	(0.0121) .	(0.0126)	(0.0127)	(0.0129) ***	(0.0126) **	(0.0132) **	(0.0133) **
90 percent or more of income from SE [3]	-0.0481	-0.0143	-0.0307	-0.0271	-0.0068	0.0217	0.0091	0.0115	-0.0048	0.0203	0.0122	0.0154
	(0.0112) ***	(0.0109)	(0.0110) *	(0.0110) .	(0.0112)	(0.0108) .	(0.0110)	(0.0110)	(0.0114)	(0.0111)	(0.0113)	(0.0113)
Male	0.0706	0.0682	0.0843	0.0844	0.0678	0.0688	0.0863	0.0860	0.0426	0.0453	0.0561	0.0560
	(0.0121) ***	(0.0117) ***	(0.0118) ***	(0.0118) ***	(0.0119) ***	(0.0115) ***	(0.0117) ***	(0.0117) ***	(0.0126) ***	(0.0122) **	(0.0124) **	(0.0124) **
Between 25 and 34	0.1038	0.0686	0.0893	0.0837	0.1440	0.1137	0.1367	0.1288	0.1415	0.1317	0.1443	0.1422
	(0.0184) ***	(0.0178) **	(0.0179) **	(0.0179) **	(0.0197) ***	(0.0191) ***	(0.0192) ***	(0.0193) ***	(0.0201) ***	(0.0195) ***	(0.0196) ***	(0.0196) ***
Between 35 and 44	0.0712	0.0390	0.0506	0.0446	0.0948	0.0671	0.0809	0.0757	0.0991	0.0862	0.0951	0.0897
	(0.0162) ***	(0.0157) .	(0.0158) *	(0.0158) *	(0.0166) ***	(0.0161) **	(0.0162) **	(0.0162) **	(0.0178) ***	(0.0172) **	(0.0173) **	(0.0173) **
Between 45 and 54	0.0783	0.0427	0.0552	0.0501	0.0627	0.0319	0.0432	0.0365	0.0536	0.0351	0.0372	0.0311
	(0.0168) ***	(0.0162) *	(0.0164) **	(0.0164) *	(0.0177) ***	(0.0171)	(0.0172) .	(0.0173) .	(0.0182) **	(0.0176) .	(0.0176) .	(0.0177)
Between 55 and 59	0.0877	0.0695	0.0799	0.0733	0.0292	0.0149	0.0198	0.0141	0.0474	0.0418	0.0441	0.0408
	(0.0142) ***	(0.0137) **	(0.0138) **	(0.0139) **	(0.0143) *	(0.0138)	(0.0139)	(0.0139)	(0.0152) **	(0.0147) *	(0.0147) *	(0.0148) *
Married filing Jointly	0.3341	0.2360	0.2328	0.2275	0.3368	0.2606	0.2681	0.2617	0.3902	0.3146	0.3164	0.3102
	(0.0223) ***	(0.0216) ***	(0.0225) ***	(0.0226) ***	(0.0223) ***	(0.0217) ***	(0.0224) ***	(0.0226) ***	(0.0248) ***	(0.0241) ***	(0.0249) ***	(0.0250) ***
Head of Household	0.0068	0.0400	0.0197	0.0310	0.0381	0.1062	0.0957	0.1073	0.1096	0.1628	0.1546	0.1724
	(0.0268)	(0.0258)	(0.0269)	(0.0271)	(0.0270)	(0.0261) **	(0.0270) **	(0.0272) **	(0.0300) ***	(0.0291) ***	(0.0300) **	(0.0302) **
1 to 3 dependents	0.0488	0.0476	0.0539	0.0542	0.0570	0.0603	0.0684	0.0707	0.0056	0.0026	0.0087	0.0106
	(0.0147) ***	(0.0141) **	(0.0142) **	(0.0143) **	(0.0153) ***	(0.0148) **	(0.0149) **	(0.0149) **	(0.0163)	(0.0158)	(0.0159)	(0.0159)
4 or more dependents	0.0519	0.0789	0.0839	0.0862	-0.0429	-0.0224	-0.0208	-0.0161	-0.0304	-0.0162	-0.0090	-0.0050
	(0.0196) **	(0.0189) **	(0.0195) **	(0.0196) **	(0.0202) *	(0.0195)	(0.0200)	(0.0201)	(0.0215)	(0.0208)	(0.0215)	(0.0216)
Using a paid preparer	0.2949	0.2743	0.3148	0.3112	0.2587	0.2155	0.2265	0.2159	0.1456	0.1113	0.1080	0.1011
	(0.0504) ***	(0.0485) ***	(0.0503) **	(0.0504) **	(0.0473) ***	(0.0456) **	(0.0472) **	(0.0473) **	(0.0493) **	(0.0478) .	(0.0505) .	(0.0507) .
Squared(Using a paid preparer)	-0.1397	-0.1499	-0.1139	-0.1055	-0.1148	-0.1030	-0.0513	-0.0391	-0.0432	-0.0282	0.0219	0.0319
	(0.0504) **	(0.0485) *	(0.0499) .	(0.0500) .	(0.0473) *	(0.0456) .	(0.0470)	(0.0471)	(0.0494)	(0.0479)	(0.0498)	(0.0499)
ρ		0.6153		0.6211		0.6068		0.5997		0.6063		0.5701
		(0.0247) ***		(0.0351) ***		(0.0251) ***		(0.0364) ***		(0.0278) ***		(0.0381) ***
λ			0.7609				0.7519				0.7211	
			(0.0264) ***				(0.0271) ***				(0.0293) ***	

Footnotes at the end of table.

Table D.3: Subgroup 2 Model Estimates, 2002-2013, cont.

	2005				2006				2007			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Lag(Total Adjusted Gross Income)				-(0.0835)				-(0.0188)				-(0.0459)
				(0.1155)				(0.1121)				(0.1130)
Lag(Squared(Total Adjusted Gross Income))				(0.1478)				(0.0577)				(0.0974)
				(0.1248)				(0.1229)				(0.1238)
Lag(Qualifying for 20 Percent Credit)				(0.0119)				-(0.1479)				-(0.2470)
				(0.0941)				(0.0942)				(0.0930) *
Lag(Qualifying for 10 Percent Credit)				-(0.0745)				(0.0176)				-(0.0467)
				(0.0577)				(0.0596)				(0.0606)
Lag(90 Percent or more of income from SE[3])				(0.0623)				-(0.0257)				-(0.0044)
				(0.0659)				(0.0634)				(0.0604)
Lag(Male)				-(0.1481)				-(0.2133)				-(0.1458)
				(0.0545) *				(0.0531) **				(0.0632) .
Lag(Between 25 and 34)				-(0.2866)				-(0.3867)				-(0.2351)
				(0.1161) .				(0.1297) *				(0.1463)
Lag(Between 35 and 44)				-(0.2019)				-(0.2196)				-(0.2940)
				(0.1026) .				(0.1033) .				(0.1234) .
Lag(Between 45 and 54)				-(0.0874)				-(0.1741)				-(0.1341)
				(0.0977)				(0.1190)				(0.1376)
Lag(Between 55 and 59)				-(0.3570)				-(0.1756)				-(0.1387)
				(0.1062) **				(0.1043)				(0.1210)
Lag(Married filing Jointly)				(0.3508)				(0.0674)				(0.1750)
				(0.1189) *				(0.1144)				(0.1370)
Lag(Head of Household)				(0.3334)				-(0.0966)				(0.1175)
				(0.1416) .				(0.1423)				(0.1617)
Lag(1 to 3 dependents)				-(0.1031)				(0.0239)				(0.0629)
				(0.0959)				(0.0971)				(0.1063)
Lag(4 or more dependents)				-(0.2386)				(0.1414)				(0.0016)
				(0.1153) .				(0.1209)				(0.1270)
Lag(Using a paid preparer)				-(0.1636)				-(0.0663)				(0.3235)
				(0.2126)				(0.1933)				(0.1962)
Lag(Squared(Using a paid preparer))				-(0.0060)				-(0.0657)				-(0.4305)
				(0.2134)				(0.1925)				(0.1973) .

Footnotes at the end of table.

Table D.3: Subgroup 2 Model Estimates, 2002-2013, cont.

	2005				2006				2007			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
N	7,455	7,455	7,455	7,455	7,453	7,453	7,453	7,453	7,476	7,476	7,476	7,476
AIC	19,732	19,224	19,172	19,140	19,717	19,241	19,197	19,163	2,092	19,676	19,675	19,633
Moran I Residuals	0.0829 ***	0.0070 **	-0.0048	-0.0028	0.0774 ***	0.0066 **	-0.0047	-0.0032	0.0653 ***	0.0001	-0.0049	-0.0023
Log Likelihood		-9,593.09	-9,566.97	-9,535.01		-9,601.50	-9,579.39	-9,546.65		-9,818.94	-9,818.64	-9,781.29
Lagrange Multiplier												
Error	1,698 ***				1,480 ***				1,061 ***			
LMlag	1,151 ***				1,026 ***				951 ***			
RLMerr	633 ***				541 ***				228 ***			
RLMlag	86 ***				87 ***				119 ***			

Footnotes at the end of table.

Table D.3: Subgroup 2 Model Estimates, 2002-2013, cont.

	2008				2009				2010			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Intercept	0.0000	0.0036	0.0153	0.0062	0.0000	0.0045	0.0181	0.0069	0.0000	0.0037	0.0128	0.0067
	(0.0106)	(0.0103)	(0.0331)	(0.0103)	(0.0108)	(0.0105)	(0.0353)	(0.0105)	(0.0108)	(0.0106)	(0.0306)	(0.0106)
Total Adjusted Gross Income	0.0268	0.0309	0.0154	0.0180	0.0103	0.0238	0.0030	0.0034	-0.0424	-0.0310	-0.0321	-0.0298
	(0.0277)	(0.0269)	(0.0286)	(0.0288)	(0.0279)	(0.0271)	(0.0288)	(0.0290)	(0.0135) **	(0.0132) .	(0.0139) .	(0.0140) .
Squared(Total Adjusted Gross Income)	-0.0218	-0.0348	-0.0246	-0.0274	-0.0101	-0.0282	-0.0129	-0.0138	-0.0050	-0.0075	-0.0061	-0.0056
	(0.0265)	(0.0258)	(0.0265)	(0.0266)	(0.0268)	(0.0260)	(0.0267)	(0.0268)	(0.0128)	(0.0125)	(0.0125)	(0.0125)
Qualifying for 20 Percent Credit	0.0689	0.0650	0.0611	0.0628	0.0224	0.0246	0.0229	0.0230	0.0031	0.0064	0.0034	0.0062
	(0.0114) ***	(0.0111) ***	(0.0111) **	(0.0111) **	(0.0121) .	(0.0117) .	(0.0117)	(0.0117) .	(0.0118)	(0.0115)	(0.0115)	(0.0116)
Qualifying for 10 Percent Credit	0.0744	0.0946	0.0917	0.0964	-0.0043	0.0180	0.0093	0.0128	0.0652	0.0839	0.0810	0.0862
	(0.0122) ***	(0.0119) ***	(0.0124) ***	(0.0125) ***	(0.0128)	(0.0125)	(0.0130)	(0.0130)	(0.0125) ***	(0.0122) ***	(0.0127) **	(0.0128) ***
90 percent or more of income from SE [3]	-0.0277	-0.0071	-0.0101	-0.0067	-0.0265	-0.0020	-0.0043	-0.0013	-0.1019	-0.0740	-0.0770	-0.0714
	(0.0111) *	(0.0108)	(0.0110)	(0.0110)	(0.0114) *	(0.0111)	(0.0113)	(0.0113)	(0.0114) ***	(0.0112) ***	(0.0115) ***	(0.0115) **
Male	0.0406	0.0389	0.0432	0.0437	0.0724	0.0750	0.0842	0.0850	0.0354	0.0383	0.0436	0.0449
	(0.0122) ***	(0.0119) *	(0.0121) **	(0.0122) **	(0.0121) ***	(0.0118) ***	(0.0120) ***	(0.0120) ***	(0.0119) **	(0.0116) **	(0.0118) **	(0.0118) **
Between 25 and 34	0.2407	0.2181	0.2327	0.2326	0.1763	0.1418	0.1566	0.1561	0.2154	0.1821	0.1897	0.1875
	(0.0207) ***	(0.0202) ***	(0.0205) ***	(0.0206) ***	(0.0203) ***	(0.0198) ***	(0.0201) ***	(0.0202) ***	(0.0188) ***	(0.0185) ***	(0.0188) ***	(0.0189) ***
Between 35 and 44	0.2361	0.2088	0.2172	0.2134	0.1990	0.1656	0.1792	0.1765	0.2319	0.2000	0.2077	0.2029
	(0.0180) ***	(0.0175) ***	(0.0177) ***	(0.0177) ***	(0.0180) ***	(0.0176) ***	(0.0178) ***	(0.0179) ***	(0.0175) ***	(0.0172) ***	(0.0174) ***	(0.0175) ***
Between 45 and 54	0.1766	0.1556	0.1552	0.1546	0.1440	0.1124	0.1177	0.1156	0.2231	0.1910	0.1962	0.1904
	(0.0189) ***	(0.0184) ***	(0.0185) ***	(0.0185) ***	(0.0187) ***	(0.0182) ***	(0.0183) ***	(0.0183) ***	(0.0184) ***	(0.0180) ***	(0.0182) ***	(0.0182) ***
Between 55 and 59	0.0857	0.0731	0.0740	0.0700	0.0222	0.0079	0.0103	0.0100	0.1372	0.1197	0.1225	0.1203
	(0.0157) ***	(0.0153) **	(0.0153) **	(0.0153) **	(0.0160)	(0.0155)	(0.0155)	(0.0156)	(0.0160) ***	(0.0157) ***	(0.0157) ***	(0.0157) ***
Married filing Jointly	0.3153	0.2530	0.2454	0.2354	0.2476	0.2016	0.2038	0.1995	0.2123	0.1551	0.1499	0.1387
	(0.0235) ***	(0.0230) ***	(0.0235) ***	(0.0236) ***	(0.0241) ***	(0.0235) ***	(0.0240) ***	(0.0240) ***	(0.0239) ***	(0.0234) ***	(0.0240) **	(0.0241) **
Head of Household	0.0665	0.1209	0.1135	0.1315	0.0422	0.1008	0.0877	0.1040	0.0124	0.0554	0.0502	0.0675
	(0.0282) *	(0.0275) **	(0.0283) **	(0.0285) **	(0.0277)	(0.0270) **	(0.0279) *	(0.0281) **	(0.0278)	(0.0272) .	(0.0280)	(0.0282) .
1 to 3 dependents	0.0584	0.0496	0.0563	0.0558	0.0617	0.0578	0.0680	0.0695	0.0520	0.0518	0.0590	0.0579
	(0.0156) ***	(0.0151) *	(0.0153) **	(0.0154) **	(0.0159) ***	(0.0155) **	(0.0157) **	(0.0157) **	(0.0163) **	(0.0159) *	(0.0161) **	(0.0162) **
4 or more dependents	-0.0115	-0.0089	-0.0064	-0.0058	-0.0569	-0.0383	-0.0217	-0.0166	-0.0308	-0.0175	-0.0111	-0.0087
	(0.0206)	(0.0200)	(0.0208)	(0.0211)	(0.0204) **	(0.0198)	(0.0206)	(0.0208)	(0.0208)	(0.0202)	(0.0210)	(0.0212)
Using a paid preparer	0.1520	0.1195	0.1383	0.1295	0.1770	0.1119	0.1101	0.0927	0.2205	0.1713	0.1747	0.1574
	(0.0513) **	(0.0499) .	(0.0526) *	(0.0529) .	(0.0501) ***	(0.0487) .	(0.0515) .	(0.0519)	(0.0490) ***	(0.0479) **	(0.0504) **	(0.0508) *
Squared(Using a paid preparer)	-0.0330	-0.0314	-0.0130	-0.0054	-0.0634	-0.0453	-0.0330	-0.0261	-0.1257	-0.1140	-0.1080	-0.1012
	(0.0514)	(0.0500)	(0.0521)	(0.0523)	(0.0502)	(0.0487)	(0.0504)	(0.0505)	(0.0490) *	(0.0478) .	(0.0494) .	(0.0495) .
ρ		0.5749		0.5232		0.6015		0.5723		0.5575		0.4629
		(0.0285) ***		(0.0410) ***		(0.0284) ***		(0.0383) ***		(0.0310) ***		(0.0444) ***
λ			0.6893				0.7028				0.6545	
			(0.0314) ***				(0.0305) ***				(0.0337) ***	

Footnotes at the end of table.

Table D.3: Subgroup 2 Model Estimates, 2002-2013, cont.

	2008				2009				2010			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Lag(Total Adjusted Gross Income)				(0.0670)				(0.0388)				-(0.0383)
				(0.1215)				(0.1160)				(0.0712)
Lag(Squared(Total Adjusted Gross Income))				(0.0089)				-(0.0443)				(0.0239)
				(0.1317)				(0.1279)				(0.0937)
Lag(Qualifying for 20 Percent Credit)				(0.1097)				(0.0088)				(0.1649)
				(0.0964)				(0.0981)				(0.0998)
Lag(Qualifying for 10 Percent Credit)				-(0.0336)				-(0.0139)				-(0.0216)
				(0.0590)				(0.0646)				(0.0636)
Lag(90 Percent or more of income from SE[3])				(0.0101)				-(0.0508)				-(0.0516)
				(0.0662)				(0.0636)				(0.0643)
Lag(Male)				-(0.0670)				-(0.1771)				-(0.0922)
				(0.0618)				(0.0615) *				(0.0641)
Lag(Between 25 and 34)				-(0.0973)				(0.0981)				(0.0389)
				(0.1343)				(0.1387)				(0.1147)
Lag(Between 35 and 44)				-(0.1891)				-(0.0720)				-(0.1140)
				(0.1171)				(0.1264)				(0.1071)
Lag(Between 45 and 54)				(0.1188)				(0.1159)				-(0.0294)
				(0.1352)				(0.1424)				(0.1290)
Lag(Between 55 and 59)				-(0.1795)				(0.0821)				-(0.0135)
				(0.1230)				(0.1282)				(0.1258)
Lag(Married filing Jointly)				(0.1905)				(0.2504)				(0.3303)
				(0.1487)				(0.1428)				(0.1318) .
Lag(Head of Household)				(0.0421)				(0.1862)				(0.1754)
				(0.1750)				(0.1572)				(0.1593)
Lag(1 to 3 dependents)				-(0.0261)				-(0.1097)				-(0.2232)
				(0.1102)				(0.1025)				(0.1035) .
Lag(4 or more dependents)				-(0.0163)				-(0.1437)				-(0.1343)
				(0.1360)				(0.1251)				(0.1324)
Lag(Using a paid preparer)				(0.0251)				(0.1716)				(0.2801)
				(0.1947)				(0.2002)				(0.2117)
Lag(Squared(Using a paid preparer))				-(0.0902)				-(0.1732)				-(0.2652)
				(0.1955)				(0.2045)				(0.2156)

Footnotes at the end of table.

Table D.3: Subgroup 2 Model Estimates, 2002-2013, cont.

	2008				2009				2010			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
N	7,449	7,449	7,449	7,449	7,438	7,438	7,438	7,438	7,424	7,424	7,424	7,424
AIC	19,833	19,477	19,495	19,458	20,108	19,732	19,741	19,722	20,063	19,757	19,796	19,754
Moran I Residuals	0.0561 ***	-0.0003	-0.0046	-0.0027	0.0603 ***	0.0014	-0.0028	-0.0017	0.0470 ***	-0.0029	-0.0035	-0.0018
Log Likelihood		-9,719.55	-9,728.64	-9,694.12		-9,847.13	-9,851.68	-9,825.92		-9,859.67	-9,878.81	-9,842.22
Lagrange Multiplier												
Error	778 ***				894 ***				541 ***			
LMlag	769 ***				829 ***				664 ***			
RLMerr	139 ***				169 ***				40 ***			
RLMlag	129 ***				104 ***				163 ***			

Footnotes at the end of table.

Table D.3: Subgroup 2 Model Estimates, 2002-2013, cont.

	2011				2012				2013			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Intercept	0.0000	0.0031	0.0090	0.0047	0.0000	0.0026	0.0086	0.0054	0.0000	0.0020	0.0080	0.0062
	(0.0108)	(0.0105)	(0.0283)	(0.0105)	(0.0110)	(0.0108)	(0.0273)	(0.0108)	(0.0111)	(0.0109)	(0.0270)	(0.0109)
Total Adjusted Gross Income	-0.0338	-0.0269	-0.0301	-0.0291	-0.0755	-0.0455	-0.0519	-0.0405	0.0156	0.0271	0.0111	0.0137
	(0.0127) **	(0.0124) .	(0.0131) .	(0.0132) .	(0.0280) **	(0.0275)	(0.0291)	(0.0295)	(0.0282)	(0.0278)	(0.0293)	(0.0296)
Squared(Total Adjusted Gross Income)	-0.0119	-0.0089	-0.0074	-0.0070	0.0228	0.0024	0.0084	0.0018	-0.0299	-0.0369	-0.0267	-0.0281
	(0.0120)	(0.0118)	(0.0118)	(0.0118)	(0.0269)	(0.0264)	(0.0270)	(0.0271)	(0.0271)	(0.0267)	(0.0272)	(0.0273)
Qualifying for 20 Percent Credit	-0.0133	-0.0112	-0.0110	-0.0116	0.0136	0.0111	0.0082	0.0097	-0.0153	-0.0178	-0.0229	-0.0216
	(0.0120)	(0.0117)	(0.0117)	(0.0117)	(0.0121)	(0.0119)	(0.0119)	(0.0119)	(0.0121)	(0.0119)	(0.0119)	(0.0119)
Qualifying for 10 Percent Credit	-0.0014	0.0095	0.0029	0.0054	0.0296	0.0325	0.0285	0.0332	-0.0091	0.0002	-0.0035	0.0021
	(0.0126)	(0.0124)	(0.0128)	(0.0128)	(0.0127) *	(0.0125) *	(0.0129) .	(0.0129) .	(0.0129)	(0.0127)	(0.0130)	(0.0131)
90 percent or more of income from SE [3]	-0.0474	-0.0247	-0.0281	-0.0242	-0.0482	-0.0330	-0.0367	-0.0303	-0.0015	0.0167	0.0148	0.0209
	(0.0114) ***	(0.0112) .	(0.0115) .	(0.0115) .	(0.0117) ***	(0.0115) *	(0.0118) *	(0.0118) .	(0.0117)	(0.0116)	(0.0119)	(0.0119)
Male	0.0184	0.0171	0.0213	0.0217	-0.0211	-0.0231	-0.0218	-0.0217	0.0391	0.0374	0.0413	0.0421
	(0.0117)	(0.0115)	(0.0116)	(0.0116)	(0.0123) .	(0.0121)	(0.0123)	(0.0123)	(0.0125) **	(0.0123) *	(0.0126) *	(0.0126) **
Between 25 and 34	0.2722	0.2570	0.2703	0.2698	0.2276	0.2110	0.2234	0.2236	0.2489	0.2369	0.2465	0.2478
	(0.0170) ***	(0.0167) ***	(0.0170) ***	(0.0170) ***	(0.0167) ***	(0.0164) ***	(0.0168) ***	(0.0169) ***	(0.0154) ***	(0.0152) ***	(0.0154) ***	(0.0155) ***
Between 35 and 44	0.2851	0.2626	0.2722	0.2688	0.2877	0.2648	0.2764	0.2728	0.2681	0.2450	0.2566	0.2522
	(0.0163) ***	(0.0160) ***	(0.0163) ***	(0.0163) ***	(0.0164) ***	(0.0162) ***	(0.0164) ***	(0.0165) ***	(0.0163) ***	(0.0162) ***	(0.0165) ***	(0.0166) ***
Between 45 and 54	0.2431	0.2257	0.2300	0.2257	0.2292	0.2068	0.2118	0.2098	0.2680	0.2458	0.2490	0.2472
	(0.0172) ***	(0.0169) ***	(0.0169) ***	(0.0170) ***	(0.0173) ***	(0.0171) ***	(0.0172) ***	(0.0172) ***	(0.0166) ***	(0.0164) ***	(0.0165) ***	(0.0166) ***
Between 55 and 59	0.1879	0.1781	0.1793	0.1764	0.1343	0.1216	0.1237	0.1216	0.1857	0.1723	0.1720	0.1706
	(0.0152) ***	(0.0149) ***	(0.0149) ***	(0.0149) ***	(0.0155) ***	(0.0152) ***	(0.0152) ***	(0.0153) ***	(0.0156) ***	(0.0153) ***	(0.0153) ***	(0.0153) ***
Married filing Jointly	0.2372	0.1903	0.1794	0.1699	0.2977	0.2573	0.2486	0.2389	0.1781	0.1481	0.1469	0.1361
	(0.0227) ***	(0.0223) ***	(0.0228) ***	(0.0229) ***	(0.0241) ***	(0.0237) ***	(0.0241) ***	(0.0241) ***	(0.0246) ***	(0.0242) ***	(0.0247) **	(0.0247) **
Head of Household	0.0196	0.0511	0.0372	0.0523	0.1486	0.1756	0.1649	0.1799	0.0344	0.0660	0.0723	0.0869
	(0.0261)	(0.0257) .	(0.0266)	(0.0269)	(0.0276) ***	(0.0271) ***	(0.0280) **	(0.0283) ***	(0.0278)	(0.0274) .	(0.0282) .	(0.0283) *
1 to 3 dependents	0.0472	0.0532	0.0647	0.0672	-0.0062	-0.0103	-0.0036	-0.0045	0.0658	0.0632	0.0696	0.0730
	(0.0163) **	(0.0160) **	(0.0162) **	(0.0163) **	(0.0166)	(0.0163)	(0.0166)	(0.0166)	(0.0166) ***	(0.0164) **	(0.0166) **	(0.0167) **
4 or more dependents	-0.0157	0.0008	0.0108	0.0158	-0.0949	-0.0936	-0.0857	-0.0831	0.0035	0.0085	0.0228	0.0352
	(0.0201)	(0.0197)	(0.0205)	(0.0208)	(0.0206) ***	(0.0202) **	(0.0211) **	(0.0214) **	(0.0211)	(0.0208)	(0.0217)	(0.0220)
Using a paid preparer	0.3185	0.2912	0.3289	0.3210	0.1228	0.0921	0.0918	0.0777	0.0694	0.0404	0.0211	-0.0046
	(0.0467) ***	(0.0458) ***	(0.0480) ***	(0.0485) ***	(0.0484) *	(0.0475)	(0.0500)	(0.0506)	(0.0494)	(0.0486)	(0.0514)	(0.0521)
Squared(Using a paid preparer)	-0.1811	-0.1863	-0.1933	-0.1894	-0.0409	-0.0370	-0.0324	-0.0311	0.0080	0.0113	0.0254	0.0328
	(0.0468) ***	(0.0458) **	(0.0472) **	(0.0474) **	(0.0485)	(0.0476)	(0.0490)	(0.0492)	(0.0496)	(0.0488)	(0.0505)	(0.0507)
ρ		0.5206		0.4760		0.5314		0.4332		0.5201		0.4015
		(0.0313) ***		(0.0439) ***		(0.0340) ***		(0.0460) ***		(0.0351) ***		(0.0476) ***
λ			0.6281				0.6036				0.5968	
			(0.0354) ***				(0.0370) ***				(0.0374) ***	

Footnotes at the end of table.

Table D.3: Subgroup 2 Model Estimates, 2002-2013, cont.

	2011				2012				2013			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
Lag(Total Adjusted Gross Income)				(0.0305)				(0.0153)				(0.1079)
				(0.0828)				(0.1205)				(0.1226)
Lag(Squared(Total Adjusted Gross Income))				-(0.0695)				-(0.0146)				-(0.1102)
				(0.1225)				(0.1358)				(0.1354)
Lag(Qualifying for 20 Percent Credit)				-(0.1241)				(0.1418)				(0.2849)
				(0.0953)				(0.1078)				(0.1001) *
Lag(Qualifying for 10 Percent Credit)				-(0.0135)				(0.0995)				(0.1262)
				(0.0695)				(0.0662)				(0.0670)
Lag(90 Percent or more of income from SE[3])				-(0.0752)				(0.0506)				-(0.0183)
				(0.0652)				(0.0649)				(0.0638)
Lag(Male)				-(0.0320)				(0.0274)				-(0.0392)
				(0.0657)				(0.0654)				(0.0642)
Lag(Between 25 and 34)				-(0.1427)				-(0.0512)				-(0.0164)
				(0.1013)				(0.1081)				(0.0943)
Lag(Between 35 and 44)				-(0.1194)				-(0.1049)				-(0.0778)
				(0.0995)				(0.0995)				(0.0956)
Lag(Between 45 and 54)				-(0.1219)				(0.1149)				(0.1588)
				(0.1218)				(0.1261)				(0.1173)
Lag(Between 55 and 59)				-(0.0872)				(0.0147)				(0.1645)
				(0.1194)				(0.1277)				(0.1293)
Lag(Married filing Jointly)				(0.3402)				(0.2366)				(0.0010)
				(0.1270) *				(0.1412)				(0.1312)
Lag(Head of Household)				(0.3200)				(0.1376)				-(0.1868)
				(0.1422) .				(0.1590)				(0.1400)
Lag(1 to 3 dependents)				-(0.2564)				-(0.0422)				(0.0112)
				(0.1035) .				(0.1032)				(0.0991)
Lag(4 or more dependents)				-(0.2594)				-(0.0273)				(0.0571)
				(0.1193) .				(0.1240)				(0.1102)
Lag(Using a paid preparer)				-(0.1672)				-(0.0219)				(0.1350)
				(0.1957)				(0.2109)				(0.2000)
Lag(Squared(Using a paid preparer))				(0.1060)				(0.0480)				-(0.0916)
				(0.1995)				(0.2184)				(0.2057)

Footnotes at the end of table.

Table D.3: Subgroup 2 Model Estimates, 2002-2013, cont.

	2011				2012				2013			
	OLS	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**	**OLS**	**SAR**	**SER**	**SDM**
N	7,397	7,397	7,397	7,397	7,373	7,373	7,373	7,373	7,355	7,355	7,355	7,355
AIC	19,850	19,593	19,602	19,578	20,137	19,921	19,945	19,912	20,137	19,959	19,968	19,921
Moran I Residuals	0.0472 ***	-0.0001	-0.0042	-0.0030	0.0385 ***	-0.0015	-0.0019	-0.0009	0.0343 ***	0.0002	-0.0017	0.0002
Log Likelihood		-9,777.64	-9,782.10	-9,754.22		-9,941.49	-9,953.71	-9,921.19		-9,960.36	-9,965.05	-9,925.34
Lagrange Multiplier												
Error	542 ***				359 ***				283 ***			
LMlag	536 ***				440 ***				329 ***			
RLMerr	87 ***				12 ***				11 ***			
RLMlag	81 ***				93 ***				57 ***			

Standard Errors are in paretheses below parameter estimates
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Standard Errors in Paretheses below parameter estimates
[1] All parameters reflect the percent for a given ZIP Code except Total AGI, which is the sum of reported AGI
[2] Lagged parameters are represented by θ in the SDM spatial model.
[3] SE - Self-Employed
Source: author's computation

Table D.4: Spatial Direct Effects, Indirect Effects, and Total Effects from the SDM for the Full Population, 2002-2013

	2002			2003			2004			2005		
Parameter [1]	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect
Making Retirement Contributions	0.9096	0.4485	1.3580	0.7913	-0.1246	0.6667	0.7466	-1.2137	-0.4671	0.6232	-0.4432	0.1800
Squared(Making Retirement Contributions)	-0.5352	0.0118	-0.5234	-0.3542	0.5574	0.2033	-0.3452	1.5218	1.1766	-0.1946	0.6425	0.4479
Total Adjusted Gross Income	-0.1078	0.2771	0.1693	-0.1774	0.2339	0.0565	-0.1648	-0.0375	-0.2023	-0.1338	-0.3269	-0.4607
Squared(Total Adjusted Gross Income)	0.0349	-0.7340	-0.6992	0.0960	-0.2407	-0.1448	0.0854	-0.1100	-0.0246	0.0610	0.3766	0.4377
Qualifying for 20 Percent Credit	0.0819	0.8359	0.9179	0.1185	1.0564	1.1750	0.1166	0.6027	0.7192	0.1070	0.4834	0.5904
Qualifying for 10 Percent Credit	0.3605	-0.0114	0.3491	0.4013	-0.0657	0.3356	0.3969	0.5103	0.9071	0.4435	0.1834	0.6269
90 Percent or more of income from SE[2]	0.0378	0.4080	0.4458	0.0488	0.3588	0.4076	0.0407	0.3711	0.4118	0.0513	0.2695	0.3207
Male	0.0019	0.0544	0.0564	0.0309	0.0621	0.0930	0.0529	-0.1812	-0.1283	0.0698	-0.0034	0.0664
Between 25 and 34	-0.0714	-0.1322	-0.2036	-0.0785	0.4266	0.3482	-0.0716	0.6195	0.5479	-0.0633	0.3845	0.3212
Between 35 and 44	-0.1065	-0.4466	-0.5530	-0.0944	-0.0534	-0.1477	-0.0799	0.0200	-0.0599	-0.0680	0.1247	0.0566
Between 45 and 54	-0.1239	0.0757	-0.0482	-0.1168	0.3532	0.2364	-0.0961	0.5320	0.4359	-0.0924	0.0970	0.0047
Between 55 and 59	-0.0739	-0.2798	-0.3537	-0.0660	0.2609	0.1949	-0.0961	0.2064	0.1103	-0.0915	0.5273	0.4358
Married filing Jointly	0.3060	1.0425	1.3486	0.2415	0.3536	0.5951	0.2928	0.0138	0.3067	0.3255	0.1796	0.5051
Head of Household	0.2880	0.1540	0.4419	0.2073	-0.4593	-0.2520	0.2836	-0.4662	-0.1825	0.3268	-0.3265	0.0004
1 to 3 dependents	-0.0661	-0.2928	-0.3590	-0.0777	0.2668	0.1891	-0.0744	0.2105	0.1360	-0.0941	0.3112	0.2172
4 or more dependents	0.0028	0.2653	0.2681	0.0538	0.7708	0.8247	0.0096	0.5599	0.5694	-0.0102	0.5736	0.5634
Using a paid preparer	-0.1466	1.4754	1.3288	-0.2131	1.3653	1.1521	-0.2500	1.1624	0.9124	-0.0975	1.0448	0.9473
Squared(Using a paid preparer)	0.2993	-1.5683	-1.2690	0.3980	-1.4918	-1.0938	0.4378	-1.3038	-0.8660	0.2782	-1.1942	-0.9160

Footnotes at the end of table.

	2006			2007			2008			2009		
Parameter [1]	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect
Making Retirement Contributions	0.6587	-0.6699	-0.0112	0.7203	-0.2792	0.4412	0.6611	-0.0192	0.6418	0.7693	-0.4453	0.3240
Squared(Making Retirement Contributions)	-0.2233	0.7424	0.5191	-0.2575	0.6251	0.3676	-0.2959	0.1450	-0.1508	-0.3510	0.5283	0.1773
Total Adjusted Gross Income	-0.1720	-0.1436	-0.3156	-0.1431	0.3854	0.2423	-0.1795	0.1350	-0.0445	-0.1744	0.3770	0.2026
Squared(Total Adjusted Gross Income)	0.0872	0.1043	0.1915	0.0706	-0.0968	-0.0263	0.0935	-0.2336	-0.1401	0.0868	-0.3969	-0.3102
Qualifying for 20 Percent Credit	0.1299	0.3374	0.4672	0.1491	0.0506	0.1997	0.0994	0.3113	0.4108	0.0921	0.3133	0.4054
Qualifying for 10 Percent Credit	0.4520	0.3320	0.7839	0.4266	0.4113	0.8378	0.4393	0.2091	0.6485	0.4167	0.2103	0.6270
90 Percent or more of income from SE[2]	0.0385	0.2126	0.2511	0.1056	0.6419	0.7475	0.0182	0.2838	0.3021	0.0223	0.3184	0.3407
Male	0.0655	-0.0768	-0.0112	0.0858	0.3441	0.4298	0.0444	0.1637	0.2081	0.0886	0.3200	0.4085
Between 25 and 34	-0.0349	0.1501	0.1151	-0.0633	-0.3229	-0.3862	0.0723	-0.0435	0.0288	0.0376	-0.7261	-0.6885
Between 35 and 44	-0.0522	0.1480	0.0958	-0.0373	-0.1009	-0.1382	0.0250	0.3544	0.3793	0.0173	-0.0125	0.0048
Between 45 and 54	-0.0674	0.0532	-0.0142	-0.0629	-0.4785	-0.5414	0.0015	-0.1082	-0.1067	0.0550	-0.3791	-0.3241
Between 55 and 59	-0.0525	0.1051	0.0526	-0.0676	0.5721	0.5045	0.0242	0.4610	0.4852	-0.0157	-0.1759	-0.1916
Married filing Jointly	0.2432	0.3471	0.5903	0.1754	-0.2450	-0.0697	0.1937	-0.1163	0.0774	0.2318	0.2015	0.4332
Head of Household	0.2123	-0.1474	0.0649	0.1543	-0.4136	-0.2594	0.1548	-0.5420	-0.3872	0.1820	-0.1489	0.0331
1 to 3 dependents	-0.0451	0.0717	0.0266	-0.0300	0.6210	0.5910	-0.0322	0.2815	0.2493	-0.0671	0.1996	0.1325
4 or more dependents	0.0564	0.1985	0.2549	0.1130	0.7960	0.9090	0.0964	0.5351	0.6315	0.1087	0.3191	0.4279
Using a paid preparer	-0.1713	0.7708	0.5995	-0.2713	1.1803	0.9089	-0.1108	0.7124	0.6016	-0.2857	0.6045	0.3188
Squared(Using a paid preparer)	0.3243	-0.9216	-0.5973	0.4201	-1.2978	-0.8777	0.2591	-0.8380	-0.5789	0.3942	-0.6592	-0.2650

Footnotes at the end of table.

	2010			2011			2012			2013		
Parameter [1]	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect
Making Retirement Contributions	0.6792	0.0933	0.7725	0.6910	-0.7118	-0.0208	0.5424	-0.7423	-0.1999	0.7073	-0.9049	-0.1976
Squared(Making Retirement Contributions)	-0.2850	0.0986	-0.1864	-0.2776	0.7899	0.5124	-0.1343	0.6926	0.5583	-0.2860	0.7881	0.5021
Total Adjusted Gross Income	-0.1711	0.3357	0.1646	-0.1638	-0.0221	-0.1858	-0.1507	0.4145	0.2638	-0.1445	0.2429	0.0984
Squared(Total Adjusted Gross Income)	0.0833	-0.3474	-0.2641	0.0770	-0.0325	0.0445	0.0623	-0.4925	-0.4301	0.0747	-0.2296	-0.1549
Qualifying for 20 Percent Credit	0.1154	0.8697	0.9851	0.1131	0.2765	0.3897	0.0851	0.1498	0.2349	0.0814	0.2691	0.3505
Qualifying for 10 Percent Credit	0.3853	-0.1774	0.2080	0.4344	0.1264	0.5607	0.4241	0.3467	0.7709	0.4197	0.3797	0.7993
90 Percent or more of income from SE[2]	0.0144	0.2770	0.2914	0.0082	0.2836	0.2918	0.0014	0.2055	0.2069	0.0265	0.0827	0.1092
Male	0.0674	0.3402	0.4076	0.0543	0.3235	0.3778	0.0375	0.2031	0.2407	0.0383	0.0229	0.0613
Between 25 and 34	0.0656	-0.3230	-0.2574	0.0107	-0.6791	-0.6684	0.0428	-0.3947	-0.3520	0.0313	-0.3246	-0.2933
Between 35 and 44	0.0199	0.1933	0.2132	-0.0105	0.2594	0.2489	0.0338	0.1550	0.1887	0.0155	0.2060	0.2215
Between 45 and 54	0.0460	-0.2414	-0.1954	0.0241	-0.3020	-0.2778	0.0286	0.1227	0.1513	0.0381	0.2663	0.3043
Between 55 and 59	0.0313	0.2697	0.3010	-0.0211	-0.1012	-0.1223	0.0213	-0.1667	-0.1455	0.0154	-0.1558	-0.1404
Married filing Jointly	0.1922	0.4073	0.5995	0.1698	0.3437	0.5135	0.2169	0.2305	0.4474	0.1819	0.5571	0.7390
Head of Household	0.1105	0.0764	0.1869	0.0520	0.2056	0.2576	0.1770	0.0554	0.2324	0.1169	0.2937	0.4106
1 to 3 dependents	-0.0072	0.1082	0.1010	-0.0179	-0.0256	-0.0435	-0.0251	0.0019	-0.0232	-0.0235	-0.3172	-0.3407
4 or more dependents	0.1697	0.0344	0.2041	0.1889	-0.0622	0.1268	0.1543	-0.0306	0.1237	0.2159	-0.3223	-0.1064
Using a paid preparer	-0.3186	0.9044	0.5858	-0.3487	0.8427	0.4940	-0.2952	1.0372	0.7420	-0.4935	1.2717	0.7782
Squared(Using a paid preparer)	0.3885	-0.9995	-0.6110	0.4262	-0.9009	-0.4747	0.3762	-1.0789	-0.7027	0.5691	-1.3061	-0.7370

[1] All parameters reflect the percent for a given ZIP Code except Total AGI, which is the sum of reported AGI
[2] SE - Self-Employed
Source: author's computation

Table D.5: Spatial Direct Effects, Indirect Effects, and Total Effects from the SDM for Subgroup 1, 2002-2013

	2002			2003			2004			2005		
Parameter [1]	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect
Making Retirement Contributions	0.6204	2.2998	2.9202	0.6919	1.0038	1.6957	0.7801	0.4430	1.2231	0.5828	0.2648	0.8476
Squared(Making Retirement Contributions)	-0.0361	-1.7298	-1.7658	-0.0599	-0.5531	-0.6130	-0.1472	0.0102	-0.1370	0.0668	0.1299	0.1967
Total Adjusted Gross Income	0.0027	0.1762	0.1788	0.0286	0.1423	0.1710	0.0323	0.0077	0.0400	0.0459	-0.1491	-0.1032
Squared(Total Adjusted Gross Income)	-0.0319	-0.4532	-0.4851	-0.0257	-0.0275	-0.0531	-0.0347	0.0369	0.0022	-0.0412	0.1969	0.1557
Qualifying for 20 Percent Credit	-0.0190	0.0732	0.0542	0.0021	-0.1643	-0.1622	-0.0066	-0.2138	-0.2204	-0.0035	-0.0794	-0.0828
Qualifying for 10 Percent Credit	0.0437	-0.3503	-0.3065	0.0155	-0.2259	-0.2105	-0.0215	-0.2984	-0.3199	0.0136	-0.2294	-0.2158
90 Percent or more of income from SE[2]	-0.0339	0.0753	0.0413	-0.0377	0.0666	0.0288	-0.0410	0.0517	0.0107	-0.0392	0.0985	0.0593
Male	0.0184	-0.0312	-0.0128	0.0169	0.1417	0.1586	0.0147	-0.0676	-0.0528	0.0334	-0.0699	-0.0365
Between 25 and 34	-0.0265	-0.0172	-0.0437	-0.0092	-0.0537	-0.0629	-0.0306	-0.0223	-0.0530	0.0120	-0.1443	-0.1323
Between 35 and 44	-0.0288	-0.0193	-0.0481	-0.0020	-0.1396	-0.1416	-0.0033	0.0512	0.0479	0.0142	0.0050	0.0191
Between 45 and 54	-0.0593	0.0761	0.0168	-0.0480	-0.1202	-0.1682	-0.0500	-0.0401	-0.0901	0.0023	-0.1188	-0.1165
Between 55 and 59	-0.0604	-0.3738	-0.4342	-0.0541	0.0913	0.0372	-0.0820	0.0452	-0.0368	-0.0208	-0.0903	-0.1111
Married filing Jointly	0.3808	0.5667	0.9475	0.3935	0.3090	0.7025	0.4340	0.2900	0.7240	0.3513	0.3862	0.7375
Head of Household	0.1023	-0.1080	-0.0058	0.0328	-0.1631	-0.1303	0.0852	-0.1040	-0.0188	-0.0116	-0.0456	-0.0573
1 to 3 dependents	0.0117	-0.3686	-0.3570	-0.0207	0.0313	0.0106	0.0228	-0.0156	0.0072	0.0281	-0.0473	-0.0192
4 or more dependents	-0.0535	-0.0600	-0.1135	-0.0505	0.3263	0.2758	-0.0476	0.1163	0.0688	-0.0043	0.1219	0.1176
Using a paid preparer	0.1941	0.3558	0.5500	0.0849	0.6380	0.7229	-0.0262	0.1164	0.0902	0.0903	0.2491	0.3394
Squared(Using a paid preparer)	-0.0631	-0.4768	-0.5400	0.0797	-0.7988	-0.7190	0.1850	-0.2928	-0.1078	0.0488	-0.3879	-0.3391

Footnotes at the end of table.

	2006			2007			2008			2009		
Parameter [1]	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect
Making Retirement Contributions	0.6103	0.8089	1.4192	0.7271	1.0073	1.7344	0.9430	0.1610	1.1040	0.8320	0.2010	1.0329
Squared(Making Retirement Contributions)	0.0603	-0.5553	-0.4951	-0.0472	-0.7075	-0.7547	-0.2863	0.0832	-0.2030	-0.1328	-0.0116	-0.1444
Total Adjusted Gross Income	-0.0022	-0.0393	-0.0415	-0.0131	-0.3984	-0.4115	-0.0501	-0.0661	-0.1162	-0.0187	-0.0642	-0.0829
Squared(Total Adjusted Gross Income)	-0.0118	0.0562	0.0443	-0.0032	0.3676	0.3643	0.0203	0.0664	0.0866	0.0034	0.0437	0.0471
Qualifying for 20 Percent Credit	0.0126	-0.1049	-0.0923	-0.0029	-0.0975	-0.1004	-0.0003	0.1297	0.1293	-0.0109	-0.1358	-0.1468
Qualifying for 10 Percent Credit	0.0130	-0.0849	-0.0719	0.0147	-0.2571	-0.2424	0.0151	0.0282	0.0433	0.0091	-0.1111	-0.1021
90 Percent or more of income from SE[2]	-0.0402	0.0459	0.0057	-0.0614	0.1523	0.0909	-0.0311	0.0572	0.0262	-0.0469	0.0928	0.0459
Male	0.0230	0.0043	0.0274	0.0164	-0.0582	-0.0418	0.0235	0.0512	0.0748	0.0319	-0.0417	-0.0098
Between 25 and 34	0.0451	-0.0474	-0.0022	0.0834	-0.0572	0.0262	0.1009	-0.1114	-0.0106	0.1019	-0.1601	-0.0582
Between 35 and 44	0.0394	0.0158	0.0552	0.0606	0.0055	0.0661	0.0811	0.1010	0.1821	0.0775	-0.0877	-0.0102
Between 45 and 54	0.0137	0.1181	0.1318	0.0463	-0.0705	-0.0242	0.0346	0.0097	0.0443	0.0879	0.0646	0.1525
Between 55 and 59	-0.0190	-0.1850	-0.2040	-0.0045	0.0732	0.0687	0.0098	0.0473	0.0571	0.0169	-0.3123	-0.2954
Married filing Jointly	0.3370	0.3956	0.7326	0.3751	0.0515	0.4266	0.3328	0.0757	0.4085	0.3192	0.3267	0.6458
Head of Household	-0.0474	0.0221	-0.0253	-0.0234	-0.2367	-0.2600	-0.0036	-0.2934	-0.2969	-0.0432	0.1100	0.0669
1 to 3 dependents	0.0595	-0.1375	-0.0780	0.0236	0.0239	0.0475	0.0472	0.0057	0.0529	0.0352	-0.0560	-0.0208
4 or more dependents	0.0264	0.0187	0.0451	-0.0048	0.1382	0.1334	0.0079	0.2286	0.2365	0.0063	0.0131	0.0194
Using a paid preparer	0.0350	0.2267	0.2617	-0.1625	0.3048	0.1423	0.0959	-0.0163	0.0796	-0.0795	0.2172	0.1376
Squared(Using a paid preparer)	0.0737	-0.3230	-0.2493	0.2314	-0.3987	-0.1673	-0.0075	-0.0591	-0.0666	0.1359	-0.2325	-0.0966

Footnotes at the end of table.

	2010			2011			2012			2013		
Parameter [1]	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect
Making Retirement Contributions	0.6140	0.3773	0.9913	0.7347	0.1756	0.9103	0.7970	-0.0681	0.7289	0.8371	0.1834	1.0205
Squared(Making Retirement Contributions)	0.0949	-0.1912	-0.0962	-0.0063	0.0237	0.0173	-0.0450	0.1139	0.0690	-0.0916	-0.1708	-0.2624
Total Adjusted Gross Income	-0.0148	-0.2191	-0.2339	-0.0411	-0.2479	-0.2890	-0.0254	0.0125	-0.0129	-0.0147	-0.0391	-0.0538
Squared(Total Adjusted Gross Income)	-0.0018	0.1790	0.1772	0.0142	0.1725	0.1866	-0.0017	-0.0205	-0.0222	0.0011	0.0758	0.0769
Qualifying for 20 Percent Credit	0.0072	0.1186	0.1258	0.0079	-0.1408	-0.1329	-0.0036	0.2370	0.2334	-0.0218	0.0853	0.0635
Qualifying for 10 Percent Credit	0.0220	-0.0904	-0.0684	0.0462	-0.2123	-0.1661	0.0194	0.0962	0.1156	0.0067	0.1135	0.1202
90 Percent or more of income from SE[2]	-0.0746	0.1300	0.0554	-0.0544	0.0606	0.0062	-0.0523	-0.0171	-0.0694	-0.0271	-0.1293	-0.1565
Male	0.0198	-0.0430	-0.0232	0.0095	-0.0896	-0.0801	-0.0140	-0.0108	-0.0248	0.0078	0.0579	0.0657
Between 25 and 34	0.0997	-0.1530	-0.0533	0.0954	-0.1430	-0.0476	0.0909	0.0051	0.0960	0.1043	0.0228	0.1271
Between 35 and 44	0.1022	0.0243	0.1265	0.1026	0.0622	0.1648	0.1050	0.0854	0.1904	0.0826	0.0563	0.1389
Between 45 and 54	0.0985	-0.1389	-0.0403	0.1046	-0.0529	0.0517	0.0897	0.2974	0.3871	0.0924	0.3219	0.4144
Between 55 and 59	0.0603	-0.1058	-0.0455	0.0480	-0.2052	-0.1572	0.0469	0.0397	0.0865	0.0561	0.0444	0.1006
Married filing Jointly	0.2749	0.3230	0.5978	0.2707	0.2571	0.5279	0.2806	0.1750	0.4555	0.3076	0.3982	0.7058
Head of Household	-0.1101	0.0769	-0.0333	-0.1165	0.0572	-0.0593	-0.0854	-0.0714	-0.1568	-0.0383	0.2628	0.2245
1 to 3 dependents	0.0439	-0.0952	-0.0513	0.0309	-0.1539	-0.1230	0.0503	-0.1093	-0.0590	0.0481	-0.2882	-0.2401
4 or more dependents	0.0234	0.0332	0.0567	0.0388	-0.0661	-0.0273	0.0284	0.0542	0.0826	0.0426	-0.1842	-0.1415
Using a paid preparer	-0.0397	0.0958	0.0561	0.0341	-0.0906	-0.0565	-0.0944	0.2761	0.1817	-0.1528	0.2524	0.0996
Squared(Using a paid preparer)	0.0810	-0.1280	-0.0471	0.0420	0.0392	0.0812	0.1423	-0.2982	-0.1559	0.2055	-0.2471	-0.0416

[1] All parameters reflect the percent for a given ZIP Code except Total AGI, which is the sum of reported AGI

[2] SE - Self-Employed

Source: author's computation

Table D.6: Spatial Direct Effects, Indirect Effects, and Total Effects from the SDM for Subgroup 6, 2002-2013

	2002			2003			2004			2005		
Parameter [1]	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect
Total Adjusted Gross Income	0.0483	0.4761	0.5245	0.0851	0.7037	0.7888	0.1010	0.8799	0.9809	0.1184	-0.0259	0.0925
Squared(Total Adjusted Gross Income)	-0.0458	-0.6465	-0.6923	-0.0541	-0.5751	-0.6292	-0.0767	-0.6143	-0.6910	-0.0805	0.2540	0.1735
Qualifying for 20 Percent Credit	0.0088	-0.3074	-0.2986	0.0035	0.2569	0.2604	0.0182	-0.0665	-0.0483	0.0063	0.0412	0.0476
Qualifying for 10 Percent Credit	0.0960	-0.2231	-0.1271	0.0266	0.0617	0.0883	0.0115	-0.0954	-0.0839	0.0252	-0.1528	-0.1276
90 Percent or more of income from SE[2]	-0.0034	0.1362	0.1327	0.0108	-0.1266	-0.1158	0.0101	0.0511	0.0612	-0.0264	0.1193	0.0929
Male	0.0788	-0.3688	-0.2900	0.0961	-0.3494	-0.2533	0.0377	-0.0744	-0.0367	0.0828	-0.2511	-0.1682
Between 25 and 34	0.0111	-0.6702	-0.6590	0.0790	-0.6391	-0.5601	-0.0007	-0.9608	-0.9615	0.0798	-0.6154	-0.5356
Between 35 and 44	-0.0258	-0.6586	-0.6844	0.0145	-0.1743	-0.1598	-0.0506	-0.3468	-0.3974	0.0418	-0.4568	-0.4150
Between 45 and 54	-0.0839	-0.3972	-0.4811	-0.0079	-0.3580	-0.3659	-0.0398	-0.2764	-0.3162	0.0492	-0.1476	-0.0984
Between 55 and 59	-0.0115	-0.8743	-0.8858	0.0096	-0.7564	-0.7468	-0.0444	-0.6820	-0.7264	0.0682	-0.8169	-0.7487
Married filing Jointly	0.3422	2.1703	2.5124	0.2474	1.4017	1.6491	0.3604	0.8785	1.2389	0.2356	1.2908	1.5263
Head of Household	0.2766	1.9412	2.2179	0.1204	0.8160	0.9364	0.2361	0.3653	0.6015	0.0368	0.9250	0.9617
1 to 3 dependents	0.0013	-0.6834	-0.6821	0.0715	-0.5525	-0.4810	0.1180	0.1565	0.2745	0.0531	-0.1822	-0.1291
4 or more dependents	0.0120	-1.2015	-1.1895	0.0533	-0.4060	-0.3527	0.0608	0.0444	0.1052	0.0832	-0.4854	-0.4022
Using a paid preparer	0.1777	0.5784	0.7561	0.4181	0.7407	1.1588	0.3580	0.0193	0.3774	0.3117	0.0778	0.3895
Squared(Using a paid preparer)	-0.0013	-0.7300	-0.7312	-0.1586	-0.9052	-1.0638	-0.1587	-0.0738	-0.2325	-0.1066	-0.1875	-0.2942

Footnotes at the end of table.

	2006			2007			2008			2009		
Parameter [1]	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect
Total Adjusted Gross Income	0.0645	0.0488	0.1133	0.0261	-0.0713	-0.0452	0.0189	0.1594	0.1784	0.0039	0.0946	0.0985
Squared(Total Adjusted Gross Income)	-0.0453	0.0752	0.0299	-0.0204	0.1966	0.1762	-0.0275	-0.0112	-0.0387	-0.0145	-0.1213	-0.1358
Qualifying for 20 Percent Credit	0.0174	-0.3382	-0.3208	0.0307	-0.5263	-0.4956	0.0645	0.2972	0.3617	0.0233	0.0510	0.0744
Qualifying for 10 Percent Credit	0.0146	0.0648	0.0794	0.0552	-0.0350	0.0202	0.0966	0.0351	0.1317	0.0127	-0.0153	-0.0027
90 Percent or more of income from SE[2]	0.0113	-0.0467	-0.0354	0.0154	0.0101	0.0255	-0.0066	0.0138	0.0071	-0.0020	-0.1199	-0.1219
Male	0.0835	-0.4017	-0.3182	0.0544	-0.2634	-0.2090	0.0432	-0.0920	-0.0488	0.0832	-0.2985	-0.2154
Between 25 and 34	0.1241	-0.7684	-0.6444	0.1400	-0.3561	-0.2161	0.2329	0.0508	0.2837	0.1587	0.4356	0.5943
Between 35 and 44	0.0730	-0.4326	-0.3596	0.0863	-0.5614	-0.4751	0.2125	-0.1615	0.0509	0.1769	0.0673	0.2442
Between 45 and 54	0.0342	-0.3779	-0.3438	0.0295	-0.2690	-0.2396	0.1571	0.4164	0.5735	0.1181	0.4231	0.5413
Between 55 and 59	0.0116	-0.4149	-0.4034	0.0392	-0.2668	-0.2276	0.0683	-0.2980	-0.2298	0.0112	0.2041	0.2153
Married filing Jointly	0.2651	0.5569	0.8220	0.3152	0.8136	1.1288	0.2392	0.6541	0.8933	0.2046	0.8473	1.0519
Head of Household	0.1068	-0.0801	0.0268	0.1755	0.4990	0.6745	0.1329	0.2313	0.3641	0.1074	0.5710	0.6784
1 to 3 dependents	0.0717	0.1647	0.2364	0.0116	0.1595	0.1710	0.0558	0.0065	0.0623	0.0685	-0.1625	-0.0940
4 or more dependents	-0.0141	0.3271	0.3130	-0.0050	-0.0029	-0.0079	-0.0061	-0.0403	-0.0463	-0.0187	-0.3559	-0.3747
Using a paid preparer	0.2168	0.1569	0.3737	0.1065	0.8813	0.9877	0.1306	0.1936	0.3242	0.0958	0.5222	0.6179
Squared(Using a paid preparer)	-0.0405	-0.2214	-0.2619	0.0260	-0.9533	-0.9273	-0.0066	-0.1940	-0.2005	-0.0287	-0.4373	-0.4660

Footnotes at the end of table.

	2010			2011			2012			2013		
Parameter [1]	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect	Direct Effect	Indirect Effect	Total Effect
Total Adjusted Gross Income	-0.0304	-0.0965	-0.1269	-0.0289	0.0315	0.0026	-0.0405	-0.0038	-0.0443	0.0146	0.1885	0.2031
Squared(Total Adjusted Gross Income)	-0.0054	0.0395	0.0341	-0.0078	-0.1382	-0.1459	0.0017	-0.0242	-0.0225	-0.0291	-0.2019	-0.2311
Qualifying for 20 Percent Credit	0.0078	0.3105	0.3184	-0.0130	-0.2461	-0.2591	0.0110	0.2563	0.2673	-0.0194	0.4593	0.4399
Qualifying for 10 Percent Credit	0.0864	0.0338	0.1202	0.0053	-0.0207	-0.0154	0.0343	0.1999	0.2341	0.0032	0.2112	0.2144
90 Percent or more of income from SE[2]	-0.0723	-0.1568	-0.2291	-0.0251	-0.1646	-0.1898	-0.0300	0.0657	0.0358	0.0209	-0.0165	0.0043
Male	0.0442	-0.1323	-0.0881	0.0215	-0.0411	-0.0196	-0.0216	0.0315	0.0099	0.0419	-0.0371	0.0048
Between 25 and 34	0.1888	0.2327	0.4215	0.2696	-0.0272	0.2425	0.2240	0.0801	0.3041	0.2484	0.1382	0.3866
Between 35 and 44	0.2027	-0.0372	0.1654	0.2689	0.0163	0.2851	0.2729	0.0233	0.2962	0.2524	0.0390	0.2914
Between 45 and 54	0.1910	0.1087	0.2997	0.2255	-0.0275	0.1981	0.2117	0.3611	0.5728	0.2493	0.4290	0.6783
Between 55 and 59	0.1207	0.0782	0.1989	0.1764	-0.0062	0.1702	0.1222	0.1183	0.2405	0.1725	0.3873	0.5598
Married filing Jointly	0.1427	0.7305	0.8732	0.1742	0.7991	0.9733	0.2420	0.5970	0.8390	0.1365	0.0924	0.2289
Head of Household	0.0695	0.3826	0.4521	0.0559	0.6546	0.7104	0.1819	0.3782	0.5601	0.0857	-0.2525	-0.1668
1 to 3 dependents	0.0559	-0.3637	-0.3078	0.0648	-0.4259	-0.3611	-0.0049	-0.0774	-0.0823	0.0733	0.0673	0.1406
4 or more dependents	-0.0101	-0.2561	-0.2661	0.0132	-0.4780	-0.4648	-0.0837	-0.1110	-0.1947	0.0357	0.1184	0.1541
Using a paid preparer	0.1609	0.6535	0.8145	0.3209	-0.0273	0.2935	0.0778	0.0207	0.0985	-0.0035	0.2214	0.2179
Squared(Using a paid preparer)	-0.1043	-0.5778	-0.6821	-0.1892	0.0302	-0.1591	-0.0308	0.0606	0.0298	0.0321	-0.1303	-0.0982

[1] All parameters reflect the percent for a given ZIP Code except Total AGI, which is the sum

[2] SE - Self-Employed

Source: author's computation

APPENDIX E – RD MODEL RESULTS WITHOUT ADJUSTMENT

I present here the results of the RD model without adjusting for bunching. Earlier I presented conclusive evidence that bunching exists and thus these results are best interpreted as the counterfactual to measuring the size of the reverse causality bias.

Table E.1 presents the effect of the credit on retirement contributions without adjusting for bunching. On average across 12 years, the credit increased retirement contributions $122.00. When broken out by rate, the 50% credit rate averaged an increase of $108.11, the 20% credit rate averaged only $30.12 increase, and the 10% rate averaged $100.51 increase. Figure E.1 shows these annual effects graphically. By credit, the 10 percent rate has the largest effect for the first 6 years. By 2009 the 50 percent rate takes over. Relative to the other rates, the 20 percent rate consistently has the smallest effect on retirement contributions. This is unsurprising given the small window of income for which this rate applies.

Table E.1: Effect of the Saver's Credit on Retirement Contributions, 2002-2013 [1]

(Amounts in $)

Year	Overall effect	Size of Credit		
		50% credit	20% credit	10% credit
2002	96.89 ***	84.22 ***	27.50 ***	83.73 ***
2003	137.10 ***	102.55 ***	32.29 ***	111.68 ***
2004	139.40 ***	102.34 ***	32.39 ***	116.37 ***
2005	147.60 ***	126.49 ***	38.73 ***	126.40 ***
2006	154.70 ***	111.54 ***	40.60 ***	133.20 ***
2007	143.00 ***	100.27 ***	38.93 **	108.00 ***
2008	133.60 ***	93.82 ***	21.18 ***	111.70 ***
2009	81.31 ***	105.11 ***	23.47 **	65.53 ***
2010	99.84 ***	95.23 **	47.71 ***	83.85 ***
2011	112.13 ***	114.58 ***	30.04 ***	90.66 ***
2012	102.74 ***	118.65 ***	15.05 ***	83.56 ***
2013	115.72 ***	142.50 ***	13.53 .	91.40 ***

[1] Bandwidth determined by employing Imbens-Kalyanamaran algorithm.

Signif. Codes: 0 '***' 0.001 '**', 0.01 '*', 0.05 '.' , 0.1 ' ', 1

Source: author's computation, Compliance Datawarehouse, IRS

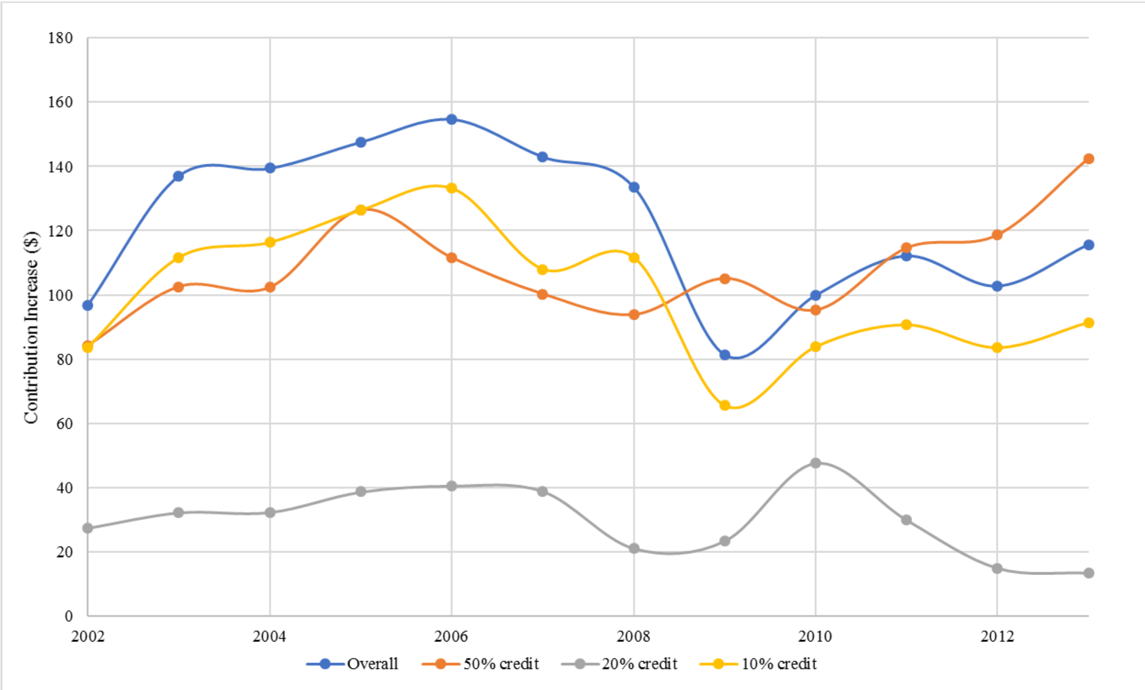


Figure E.1: Effect of the Saver's Credit on Retirement Contributions, 2002-2013

APPENDIX F – SAVER'S CREDIT BEHAVIOR DISAGGREGATED BY INCOME SOURCE

Here I disaggregate bunching between self-employed and wage earners. There is a large body of research discussing the differing ability to, and observance of, income manipulation divided between these two income sources.[31] Intuitively, the limited amount of third-party reporting, and thus external oversight, on self-employed sources of income to the IRS leads to an incentive by these taxpayers to and selectively choose to (under)report income tin order to more perfectly optimize tax benefits. For this division I define wage earners as having 90 percent or more of income coming from wages and salaries, represented by WAGE_TP1 equaling 1, and define self-employed as all other taxpayers, represented by WAGE_TP1 equaling 0.

The reason for this deviation from the definition of self-employed used above is twofold. First, even a small fraction of self-employment can be used to adjust where exactly a taxpayer falls between having income low enough to qualifying for the credit verse not, thus constraining the income source to being all or nearly all coming from wages limits the possibility of income manipulation. Second, and perhaps merely the other side of the same coin, there is no floor to the fraction of how much income can(not) come from self-employment and thus employing a minimum threshold of 90 percent of income coming from self-employment, as used above, arbitrarily excludes many taxpayers capable and incentivized to manipulate income.

[31] See Andreoni, Erard, and Feinstein (1998), Slemrod (2007), Morse et al. (2009), Saez (2010), Chetty, Friedman, and Saez (2013), and IRS (2019) among others.

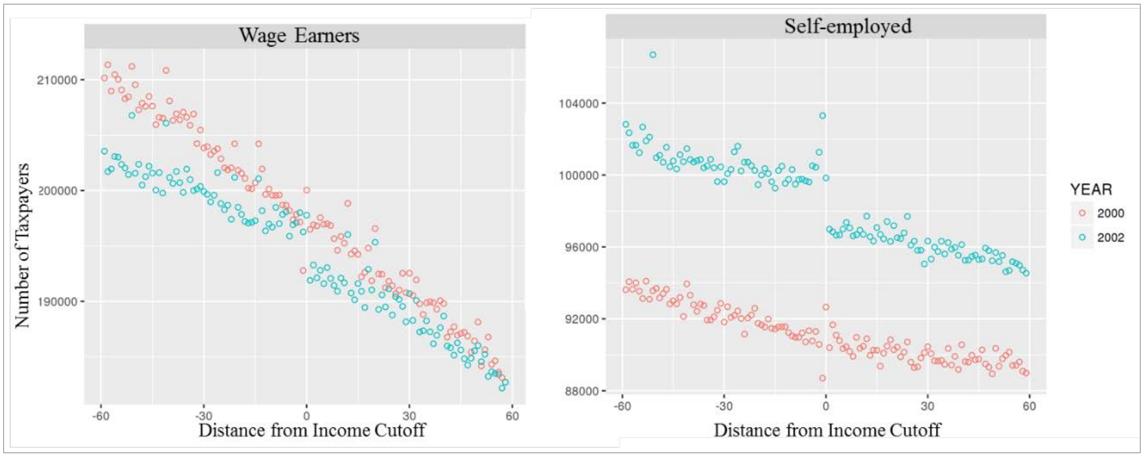


Figure F.1: Density Distribution for the Number of Taxpayers around the Saver's Credit Income Cutoff by Income Source, 2000 and 2002

Figure F.1 presents the density plot for the number of taxpayers around the AGI notch-point for years 2000 and 2002 distributed by wage earners (right) and self-employed (left). While the exact number of taxpayers represented on the vertical axes differ between the two, with wage earners representing far more taxpayers than not, we still see a pattern of discontinuity for 2002 that is not present in 2000. Upon visual inspection it appears the discontinuity present for self-employed taxpayers is slightly larger than the wage earners counterpart. This is consistent with the expectation that self-employed are more capable and incentivized to bunch around the notch-point. When expanding to the full set of tax years post credit introduction, Figure F.2, looking at the red portion of the figure, we see more indications of discontinuity than the teal counterpart, though there too we see some suggestion of discontinuity.

Following the approach discussed in Section 7.1.1, we can quantify the exact magnitude of discontinuity at the notch-point to better determine whether taxpayers are shifting income in order to qualify. Table F.1 presents the estimated density discontinuity

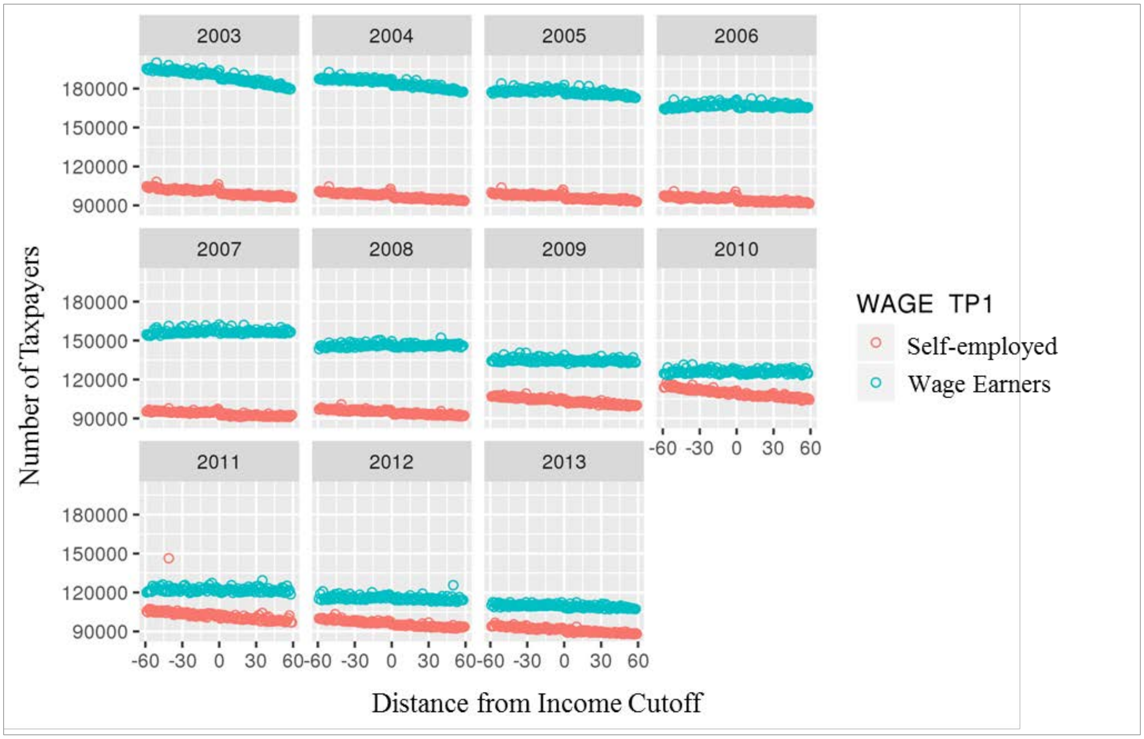


Figure F.2: Density Distribution for the Number of Taxpayers around the Saver's Credit Income Cutoff by Income Source, 2003-2013

Table F.1: Estimated Number of Excess Taxpayers Below the Income Cutoff by Source of Income, 1999-2013

	Income Source	
Year	Wage Earner	Self-Employed
1999	-390	-914 .
2000	-1,695 *	-1,163 .
2002	3,007 ***	3,703 ***
2003	2,190 *	3,917 ***
2004	2,505 **	3,510 ***
2005	2,014 *	3,629 ***
2006	1,978 *	3,215 **
2007	769	2,504 ***
2008	1,198	2,142 ***
2009	-27	987 .
2010	1,120 *	490
2011	1,961 *	1,185 .
2012	2,269 ***	2,420 ***
2013	1,732 **	1,518 ***

[1] Bandwidth is determined by employing Imbens-Kalyanamaran algorithm.

Signif. Codes: 0 '***' 0.001 '**', 0.01 '*', 0.05 '.' , 0.1 ' ', 1 for Tax Years 1999 through 2013 for wage earners and self-employed. Values greater than zero represent the estimated additional number of taxpayers above what would be expected if the density of reported AGI was continuous across the credit qualifying notch-point while negative values indicate fewer. Across the two income sources we see both subpopulations are consistently bunching just below the notch point. The most notable deviations arise in years 2007 through 2010. When comparing between income sources, Figure F.3, the first 8 years of the credit's existence shows larger magnitudes of bunching for self-employed than wage earners. This consistent trend ends in 2010, the exact explanation for which is yet unclear.

Figure F.3: Estimated Number of Excess Taxpayers Below the Income Cutoff by Source of Income, 1999-2013

APPENDIX G – BUNCHING AT EACH CREDIT RATE

Above, I employed two methods to test the RD assumption of continuous density distribution of AGI around the notch-point. First, I examined the density discontinuity precisely at the notch-point. Then, I expanded the scope to include a larger bunching region, assuming that taxpayers are not always capable of exactly controlling their reported AGI. Above, I aggregated the distribution across the various credit rate thresholds, i.e. notch-points, and in so doing obfuscated the distinction between manipulating income to qualify for claiming a Saver's Credit benefit and manipulating income to qualify for a better Saver's Credit benefit. This Appendix examines each credit rate separately.

Much like above, this Appendix is broken out into two subsections. The first section examines density discontinuity at the notch-point for each credit rate. The second section employs the Chetty et al. (2011) method for estimating bunching using the same method as presented in the paper.

Section G.1 – Density Discontinuity Approach

Table G.1 presents the estimated density discontinuity around the 50 percent credit rate cutoff for Tax Years 1999 through 2013. Values greater than zero suggest a bunching to the left of the notch-point and therefore qualifying for the Saver's Credit. While most years indicate statistical significance, years 2007 and 2008 do not. Comparing these results to the aggregated trend, where 2009 initially presented with no significance, we see here that 2009 now shows significant bunching. In fact, prior to

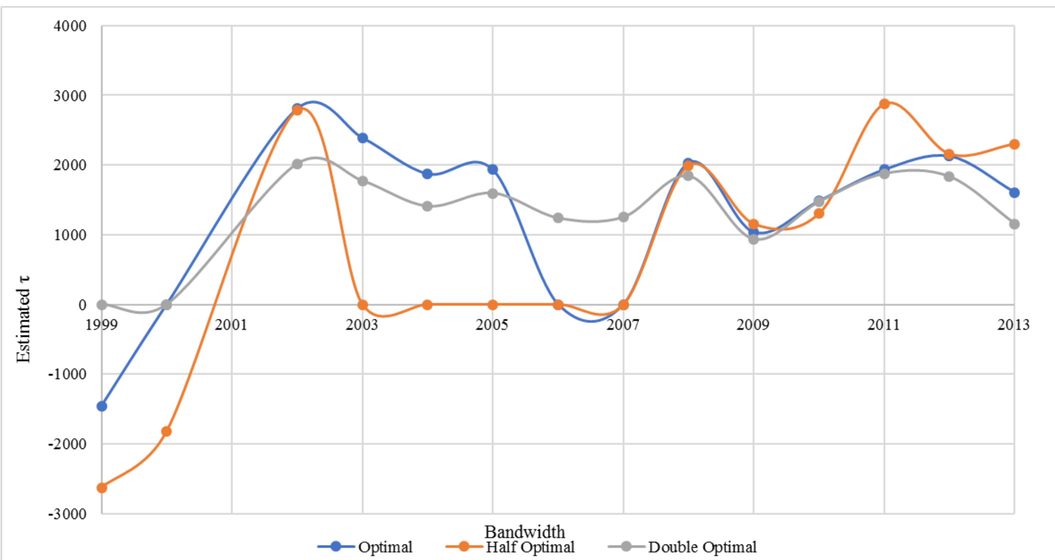
Table G.1: Number of Excess Taxpayers Below the 50% Credit Rate Income Cutoff, 1999-2013

	Size of Bandwidth [1]		
Year	Optimal	Half Optimal	Double Optimal
1999	-1,461 *	-2,629 **	-419
2000	-813	-1,819 .	-621
2002	2,811 *	2,787 .	2,016 **
2003	2,391 *	2,839	1,772 **
2004	1,873 *	2,201	1,407 *
2005	1,932 .	1,981	1,596 *
2006	1,354	1,475	1,238 .
2007	1,169	1,102	1,251 *
2008	2,033 ***	1,995 *	1,841 ***
2009	1,036 *	1,155 *	934 *
2010	1,486 ***	1,306 **	1,472 ***
2011	1,936 *	2,879 ***	1,875 ***
2012	2,128 ***	2,159 ***	1,831 ***
2013	1,609 ***	2,304 ***	1,156 **

[1] Bandwidth is determined by employing Imbens-Kalyanamaran algorithm.

Signif. Codes: 0 '***' 0.001 '**', 0.01 '*', 0.05 '.' , 0.1 ' ', 1

Source: author's computation, Compliance Datawarehouse, IRS



Note: Only statistically significant estimates are reported, all others are reported as zero

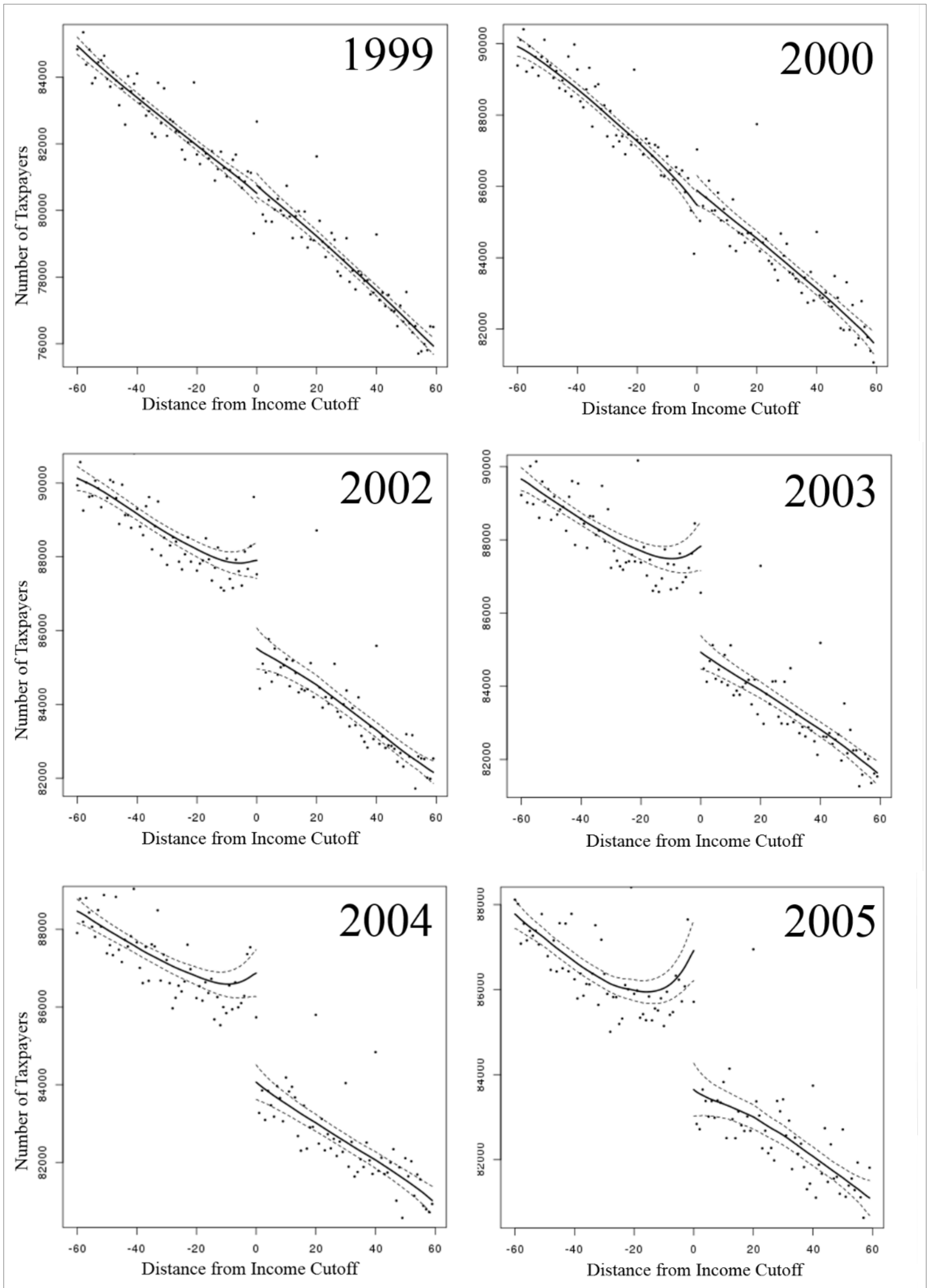
Figure G.1: Number of Excess Taxpayers Below the 50% Credit Rate Income Cutoff, 1999-2013

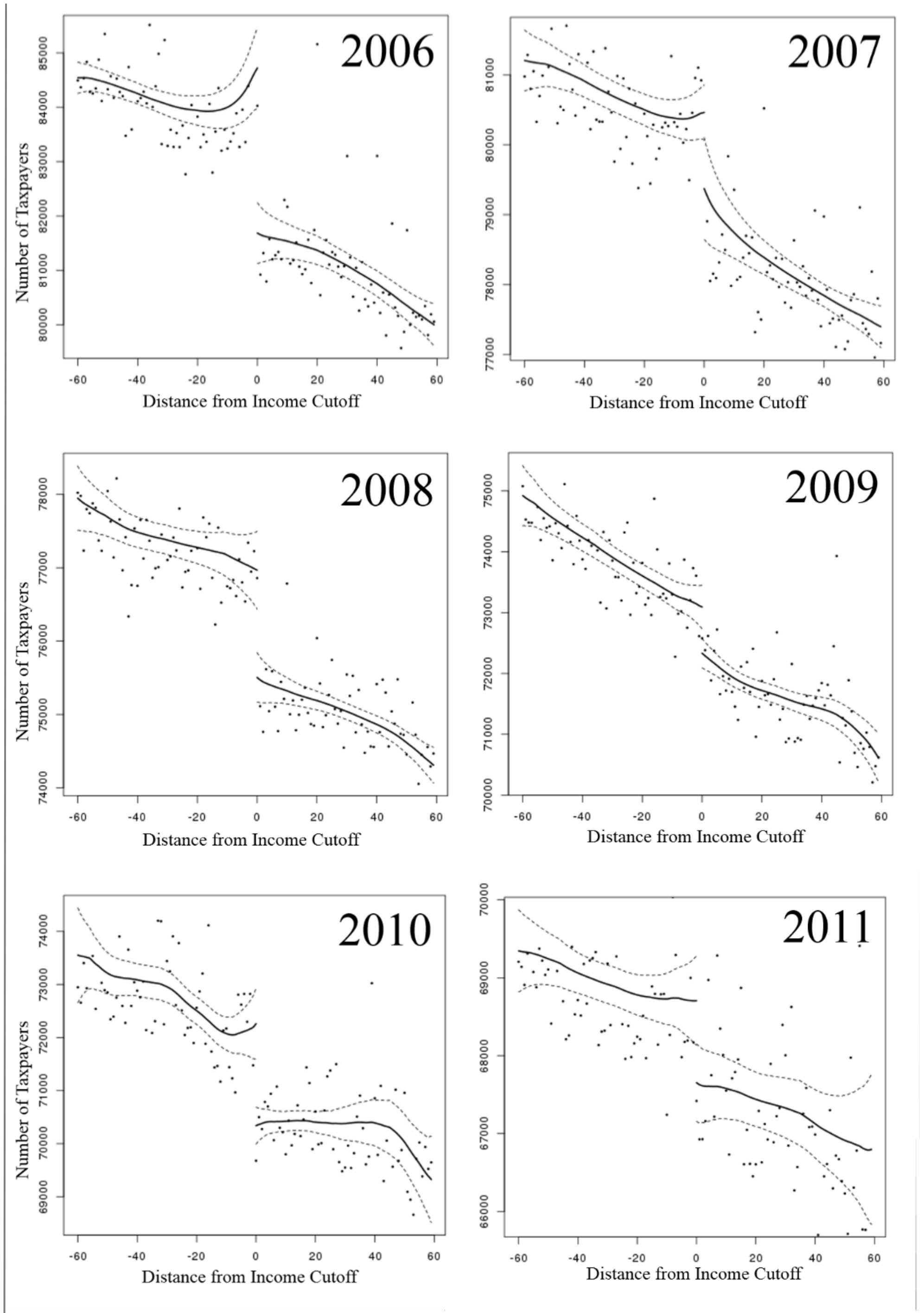
2008, for the 50 percent credit rate, the observance of discontinuity is largely based on the bandwidth chosen. If using half the optimal bandwidth, it is only in 2002 and then not until 2008 that discontinuity is estimated.

Figure G.1 visually presents these results. Much like the estimates from Chapter 7 above, there is no bunching prior to the introduction of the Saver's Credit in 2002. An estimated 2,811 more taxpayers were left of the notch-point in that first year. Subsequently bunching estimates declined through 2007. Beginning in 2008, bunching appears to have begun picking up again through 2013.

While the computed statistical significance would suggest 2006 and 2007 did not have discontinuity in the density distribution, when observing each year's density distribution, there is reason to believe that some bunching at the notch-point may have occurred. Figure G.2 shows the annual distribution and overlays a local linear regression following the same method as above. Tax Years 2005 and 2008, both statistically significant, have the same overall distribution as 2006 and 2007.

Recall from Table 1.1, the window for qualifying for the 20 percent rate is quite narrow, with an income spread of only $2,000 to $3,000, making it much harder for a taxpayer to target this precise window. Table G.2 presents the estimated density discontinuity around the 20 percent credit rate cutoff for Tax Years 1999 through 2013. Unlike any other density discontinuity estimated, virtually no estimate is statistically


1999
2000
2002
2003
2004
2005
Number of Taxpayers
Distance from Income Cutoff


2006
2007
2008
2009
2010
2011
Number of Taxpayers
Distance from Income Cutoff

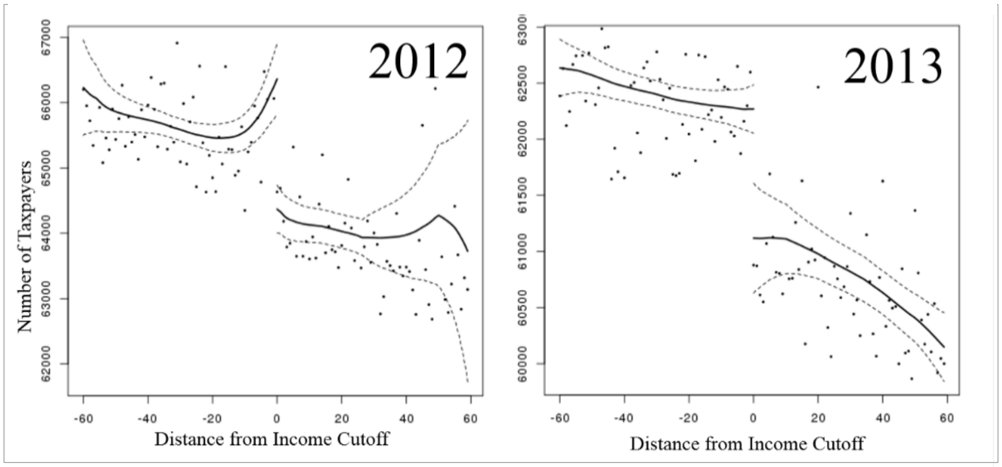


Figure G.2: Density Discontinuity Around the 50% Credit Rate, 1999-2013

Table G.2: Number of Excess Taxpayers Below the 20% Credit Rate Income Cutoff, 1999-2013

	Size of Bandwidth [1]		
Year	Optimal	Half Optimal	Double Optimal
1999	102	-617	-80
2000	-1,487 **	-1,848 *	-706 *
2002	1,227 ***	1,202 *	1,218 ***
2003	359	-230	263
2004	614	474	281
2005	440	372	438
2006	615	716	667 *
2007	723 .	772	646 .
2008	-460	-488	-443
2009	-790	-1,767	-702
2010	-1,068	-1,243	-1,318 *
2011	-660	-1,092	-1,349 *
2012	617	459	61
2013	317	206	33

[1] Bandwidth is determined by employing Imbens-Kalyanamaran algorithm.

Signif. Codes: 0 '***' 0.001 '**', 0.01 '*', 0.05 '.' , 0.1 ' ', 1

Source: author's computation, Compliance Datawarehouse, IRS

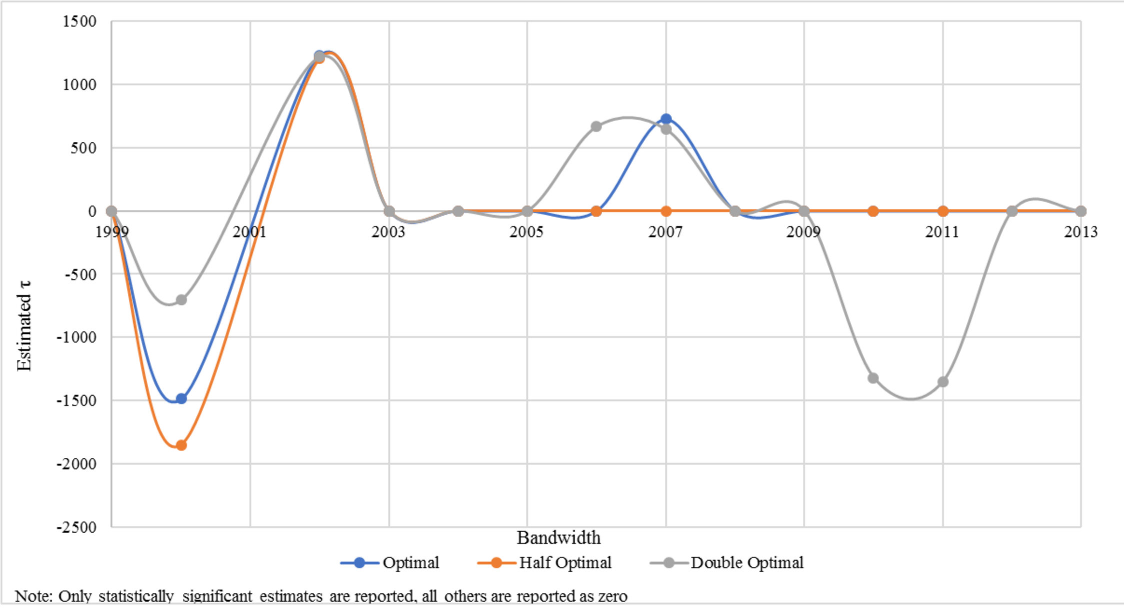


Figure G.3: Number of Excess Taxpayers Below the 20% Credit Rate Income Cutoff, 1999-2013

significant. The only years for which there is measurable discontinuity include 2002 and 2007.

Figure G.3 presents the annual distribution of excess taxpayers around the 20 percent notch-point. There is very little variation in the distribution for 1999 through 2003. In 2004 the distribution of taxpayers begins to deviate from its previously smooth pattern and by 2006 the distribution is wildly disparate until smoothing out again in 2013. The reason for these large variations is due to the combination of a taxpayer's inability to precisely manipulate income to, the small window for which the 20 percent credit rate applies and Earned Income Tax Credit income cutoffs falling within these general limits.[32]

[32] For 2002, the AGI limits were $11,060 ($12,060 for married filing jointly) for households with no qualifying child in residence, and $33,178 ($34,178 for married filing jointly) for households with a qualifying child in residence.

Turning to the 10 percent credit rate, Table G.3 and Figure G.4, the pattern observed in Chapter 7 is quite apparent here. The first five years of the credit, when qualifying income was not indexed for inflation, we observe consistently large estimated effects. After indexing, the size of bunching is dampened but remains present. In fact, looking strictly at years after 2006, there appears to be a subtle and oscillating increasing trend. More years of data would be required to confirm this however. From these results we can conclude that the observed bunching in Chapter 7 are being largely driven by income manipulation to qualify for the 10 percent credit rate. This is perhaps unsurprising given the already determined positive link between income and retirement contribution behavior and the fact that the 10 percent credit rate is for the highest income category.

Table G.3: Number of Excess Taxpayers Below the 10% Credit Rate Income Cutoff, 1999-2013

	Size of Bandwidth [1]		
Year	Optimal	Half Optimal	Double Optimal
1999	-240	-365	-316
2000	-415	-1,024 .	-250
2002	2,386 ***	2,843 ***	1,855 ***
2003	2,896 ***	3,755 ***	2,246 ***
2004	2,808 ***	3,767 ***	2,344 ***
2005	3,269 ***	4,112 ***	2,397 ***
2006	3,037 ***	3,943 ***	2,191 ***
2007	1,090	768	1,115 *
2008	1,462 ***	1,269 *	1,533 ***
2009	760 **	637 .	866 ***
2010	1,921 ***	2,187 **	1,386 ***
2011	1,052 *	1,155 *	857 *
2012	1,992 ***	2,334 ***	1,417 ***
2013	1,150 ***	1,416 ***	977 ***

[1] Bandwidth is determined by employing Imbens-Kalyanamaran algorithm.

Signif. Codes: 0 '***' 0.001 '**', 0.01 '*', 0.05 '.' , 0.1 ' ', 1

Source: author's computation, Compliance Datawarehouse, IRS

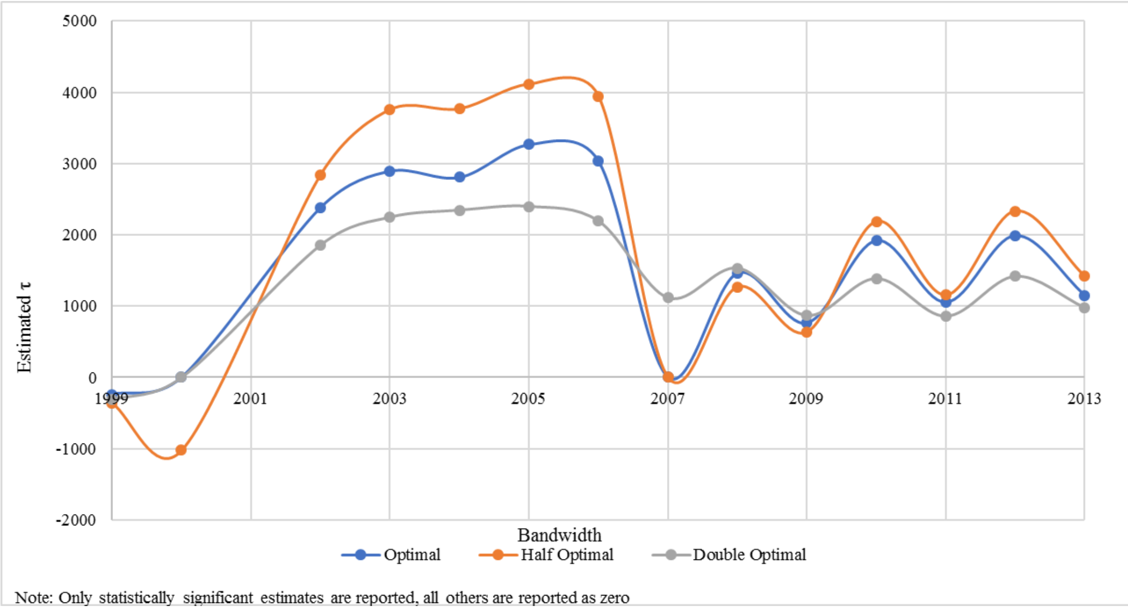


Figure G.4: Number of Excess Taxpayers Below the 10% Credit Rate Income Cutoff, 1999-2013

Section G.2 – Bunching in Density Region Approach

While the observance of discontinuity in the density is proof enough of an RD assumption violation, it does not capture the full magnitude of bunching that is occurring. The method of producing a counterfactual from which to compare the actual density distribution, described in more detail above, is a much more robust way of capturing the true size of taxpayer bunching. While I presented the total size of bunching above, by aggregating across all credit rates, here I isolate each rate independently so as to better understand which rate accounts for the largest (smallest) bunching behavior using this counterfactual approach.

Table G.4 presents the results of this approach for each credit rate and are summarized in Figure G.5. The results here are in some conflict with results from the

Table G.4: Estimated Number of Excess Taxpayers Above Counterfactual by Credit Rate Income Cutoff, 1999-2013 [1]

	50% credit rate		20% credit rate		10% credit rate	
Year	Number of Taxpayers	Percent of all Taxpayers	Number of Taxpayers	Percent of all Taxpayers	Number of Taxpayers	Percent of all Taxpayers
1999	-778 ***	-0.032 ***	7,113 ***	0.184 ***	2,812 ***	0.074 ***
2000	6,068 ***	0.169 ***	-151 ***	-0.005 ***	-1,633 ***	-0.054 ***
2002	17,385 ***	0.571 ***	-22,553 ***	-0.434 ***	9,902 ***	0.328 ***
2003	28,644 ***	0.792 ***	14,083 ***	0.358 ***	13,199 ***	0.416 ***
2004	14,740 ***	0.424 ***	5,757 ***	0.159 ***	14,201 ***	0.452 ***
2005	34,202 ***	0.762 ***	9,538 ***	0.273 ***	17,067 ***	0.746 ***
2006	42,527 ***	0.860 ***	13,141 ***	0.419 ***	14,310 ***	0.494 ***
2007	18,396 ***	0.646 ***	-764 ***	-0.022 ***	11,189 ***	0.343 ***
2008	37,983 ***	0.936 ***	814 ***	0.025 ***	7,354 ***	0.276 ***
2009	1,317 ***	0.048 ***	5,500 ***	0.203 ***	4,917 ***	0.218 ***
2010	2,736 **	0.090 **	-4,191 **	-0.147 **	3,828 **	0.256 **
2011	49,318 ***	1.147 ***	5,528 ***	0.264 ***	11,631 ***	0.438 ***
2012	13,928 ***	0.569 ***	11,197 ***	0.569 ***	9,647 ***	0.595 ***
2013	7,709 ***	0.268 ***	4,367 ***	0.192 ***	10,327 ***	0.408 ***

[1] Bandwidth is determined by employing Imbens-Kalyanamaran algorithm.
Significance based on bootstrapped standard errors from 10,000 iterations.
Signif. Codes: 0 '***' 0.001 '**', 0.01 '*', 0.05 '.' , 0.1 ' ', 1

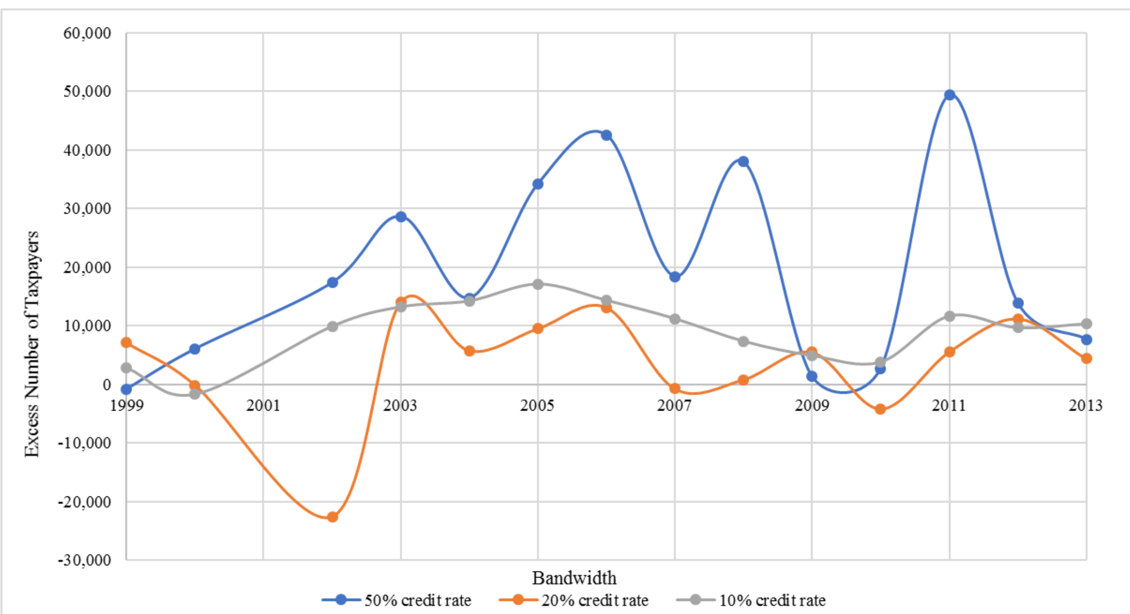


Figure G.5: Number of Excess Taxpayers Below the 10% Credit Rate Income Cutoff, 1999-2013

density discontinuity results. Where before the 10 percent rate presented with the largest estimates of bunching, here the 50 percent credit shows the largest incidence of bunching. Though there is a wide range of estimates, ranging from only 1,317 taxpayers in 2009 to 49,318 in 2011, there are only a handful of years in which the bunching magnitude for the 50 percent rate is overshadowed by another rate, and in those particular years the magnitude of bunching overall was quite low. This suggests that the 50 percent rate, not the 10 percent rate suggested before, is the driving force behind the overall magnitude of bunching presented above. What these seemingly conflicting results suggest is that those falling around the 50 percent notch-point are either less capable of precisely manipulating their income to fall just at or below the notch-point or are also attempting to optimize other low-income tax credits such as the Earned Income Tax Credit.

The 20 percent rate has the least estimated amount of bunching of all three rates. Such findings are consistent with both the density distribution analysis from Section G.1 and with the impact analysis presented in the main text. This serves to further highlight the limited impact the 20 percent rate has on taxpayer behavior.

While the 50 percent rate may be the largest in magnitude, the 10 percent rate shows the most consistent bunching magnitudes over time. The range of estimated bunching for years in which the Saver's Credit was available goes from 3,828 in 2010 to 17,067 in 2005. The drop in 2010 corresponds with the drop in all rate estimates. However, the drop experienced for this credit rate was far smaller than the drop experienced by the other two credit rates.

APPENDIX H – THE DISCREPANCY IN BUNCHING ESTIMATES FOR 2010 AND 2013

Figure H.1 presents the density distribution for 2010 with the computed counterfactual regression model overlaid. The blue dots represent observations that were used in the production of the counterfactual model, using a 5th-degree polynomial. The red dots represent the observations removed. The green dots represent the predicted density estimates. Here, more of the observed density falls below the computed counterfactual. Comparing the red dots just left and just right of the notch-point, it is visually easy to see a small discontinuity. Further away from that notch-point however, the observed density points quickly fall below the counterfactual line. When disaggregated by credit rate, see Appendix G, it becomes clear that the 50 percent and 10 percent rate both present with positive estimates of bunching but are counterbalanced by the 20 percent rate reporting no bunching.

Figure H.2 presents the density distribution for 2013 with the computed counterfactual regression model overlaid. Again, the blue dots represent observations that were used in the production of the counterfactual model, using a 5th-degree polynomial. The red dots represent the observations removed. The green dots represent the predicted density estimates. Here, there appears a generally strong discontinuity in the distribution, however, the variance in the distribution is rather large and serves to dampen the overall estimated bunching effect.

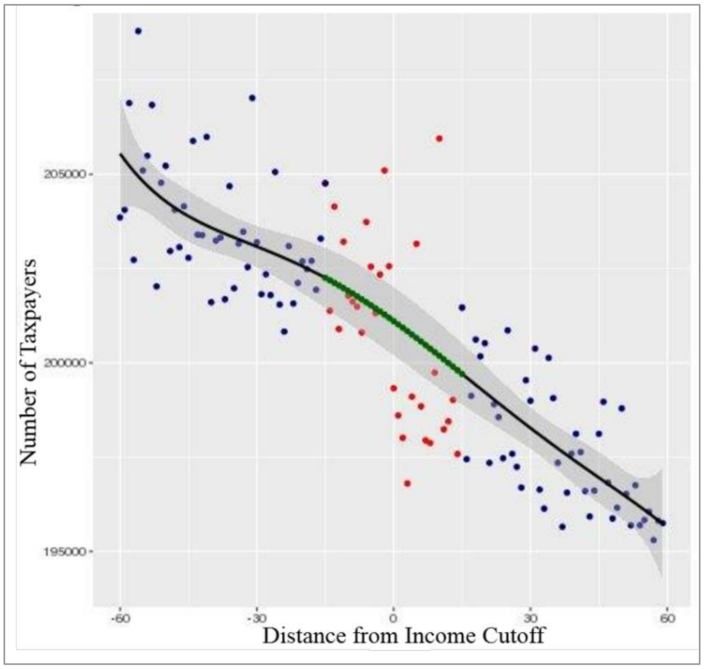


Figure H.1: Actual and Counterfactual Distribution of Taxpayers, 2010

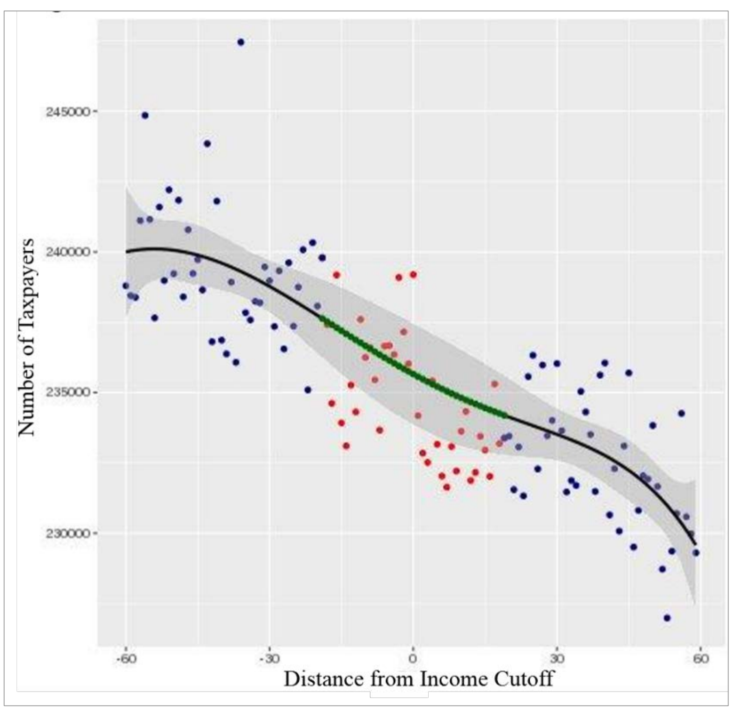


Figure H.2: Actual and Counterfactual Distribution of Taxpayers, 2013

APPENDIX I – PRIOR YEAR INCOME PROXY APPROACH FOR ACCOUNTING FOR INCOME MANIPULATION

The approach used above to adjust for income manipulation rested on accepting two assumptions surrounding the behavior of taxpayers and could only provide a range for the treatment effect. A different approach for correcting income manipulation is to use reported income prior to the introduction of the Saver's Credit as the running variable in the RD design. The intuition here is that, while taxpayers may have manipulated their income in the year for which the credit existed, those same taxpayers did not manipulate their income in years prior to the existence of the credit, as there was no incentive at the time to do so. As such, in this section I utilize that fact by using taxpayers' reported income in 2001 as the running variable.[33]

This approach affords two significant advantages over the previous adjustment approach. First, the previous approach required relying on assumptions regarding how individuals respond to the incentive to manipulate income. In actuality those choosing to manipulate income were not necessarily the lowest or the highest contributors to retirement programs but located throughout the distribution of contributions. The benefit of using prior year income as the running variable, is that it need not make any contribution assumptions. And as such, this method allows for a more precise estimate of

[33] All previous research here has entirely omitted 2001 as a study year for reasons of incomplete information surrounding 401(k) contributions. However, since I rely exclusively on reported AGI, which was not affected by the information loss, it is acceptable and quite necessary to bring back 2001 into this analysis.

Table I.1: Annual Income Correlation with 2001 Income, 2002-2004

	Year		
Correlation method	2002	2003	2004
Pearson	0.5503	0.4231	0.4356
Spearman Rank	0.7386	0.6395	0.5715

the treatment effect. In other words, as opposed to producing estimates that suggest a range of probable treatment effects, this method produces a precise, single point estimate.

The second advantage of this approach is the ability to retain all taxpayers in the analysis. In the previous approach I removed taxpayers based on their location along the retirement contribution distribution. In this method, that requirement disappears. In so doing, I am able to retain those taxpayers who may have, in response to the Saver's Credit, made large (small) contributions that I omitted from the analysis in the prior approach.

There is one fundamental disadvantage; prior year income is not perfectly correlated with current income and that imperfect correlation worsens over time. Income reported in 2001 may be to some extent a reasonable proxy for income in 2002, but as time passes that allowance breaks down. It is entirely unreasonable to assume income reported in 2001 is highly correlated with income reported in, say, 2013. Table I.1 shows the correlation between 2001 income and income reported in 2002-2004. In just one year, the linear correlation is only 0.5503. By 2004 the correlation drops to 0.436. Assuming instead that the relationship of income between years is nonlinear, the correlation improves to 0.739 in 2002, but by 2004 this drops to 0.571, just slightly better than the

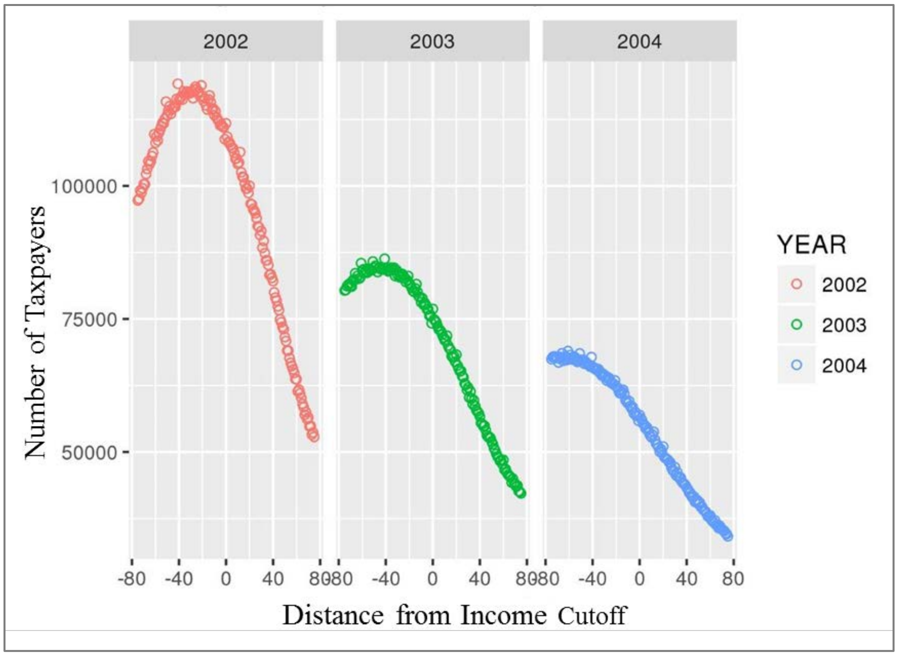


Figure I.1: Density Distribution Using 2001 Income, 2002-2004

Pearson estimate. Given that by 2004 the correlation estimates are so low, I only use this method to estimate the impact for these three years.

Before running the RD model, it is important to first ensure the running variable is smooth lest this approach offer no tangible improvements to the one prior. As expected, the density distribution for these years appears smooth across the notch-point, Figure I.1. But, as before, this visual assessment is not sufficient to conclude that bunching is not occurring. Table I.2 reports the results from running the same density discontinuity method outlined in Section 7.1.1 both overall and broken out by credit size. These results confirm the absence of bunching suggested from Figure I.2. Given these complimentary results it is reasonable to assume that the RD assumption of a smooth running variable is being satisfied.

Table I.2: Number of Excess Taxpayers Below Income Cutoff, 2002-2004

Year	Overall	Size of Credit 50%	20%	10%
2002	-227.2	-205	204	-40
2003	-419.8	-440 *	71	-14
2004	-100.8 .	-288 .	114 .	64 .

[1] Bandwidth is determined by employing Imbens-Kalyanamaran algorithm.

Signif. Codes: 0 '***' 0.001 '**', 0.01 '*', 0.05 '.' , 0.1 ' ', 1

Turning to estimating the impact, Table I.3 presents the results. Results here suggest very little, and in most cases no, impact from the Saver's Credit on retirement contributions. Even when found to be statistically significantly different from zero, the estimate is not economically significant, never reaching more than $10.00. Using this approach, it would seem the Saver's Credit has had no effect on retirement contributions.

These results are in direct contradiction to Section 7.2's results. The most likely explanation for this inconsistency lies in the lackluster correlation of income between years. imperfection of using prior year income to proxy for current year income. In order to reconcile these differences, one would need to correct for this weak instrumental variable or select an entirely different variable to proxy on.

Table I.3: Estimated Effect of Saver's Credit on Retirement Contributions Using 2001 Income, 2002-2004

(Amounts in $)

Year	Overall effect	Size of Credit 50% credit	20% credit	10% credit
2002	2.40	2.79	6.11 *	-2.11
2003	5.28 *	7.27 .	1.40	-4.16
2004	4.13	9.61 .	-0.87	-2.68

[1] Bandwidth determined by employing Imbens-Kalyanamaran

Signif. Codes: 0 '***' 0.001 '**', 0.01 '*', 0.05 '.' , 0.1 ' ', 1

WORK CITED

Adhikari, B., Harris, T., and Alm, J. (2019) "Taxpayer responses to third-party income reporting: Evidence from spatial variation across the U.S." Presented at the 9th Annual IRS-TPC Joint Research Conference on Tax Administration, Washington D.C. June 20, 2019.

Albouy, D. 2009. The unequal geographic burden of federal taxation. *Journal of Political Economy*. 117(41): 635-667.

Alm, J., Bloomquist, K., McKee, M. 2013. When you know your neighbor pays taxes: Information, peer effects, and tax compliance. Appalachian State University Working Paper No. 13-22.

Alm, J., Hawley, Z., Lee, J.M., and Miller, J. 2016. Property tax delinquency and its spillover effects on nearby properties. *Regional Science and Urban Economics.* 56: 71-77.

Alm, J, and Yunus, M. 2009. Spatiality and persistence in US individual income tax compliance. *National Tax Journal* 62(1): 101-124.

Andreoni, J., Erard, B., and Feinstein, J. 1998. Tax compliance. *Journal of Economic Literature* 36(2): 818-860.

Anselin, L. 1988. Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geographical Analysis*. 20(1): 1-17.

Anselin, L. 1995. Local Indicators of Spatial Association – LISA. *Geographic Analysis,* 24(7): 93-115.

Alatas, V., Banerjee, A., Chandrasekhar, A., Hanna, R., and Olken, B. 2016. Network Structure and the aggregation of information: Theory and evidence from Indonesia. *American Economic Review* 106(7): 1663-1704.

Baicker, K. 2005. The spillover effects of state spending. *Journal of Public Economics.* 89:529-544.

Baskaran, T. 2014. Identifying local tax mimicking with administrative borders and a policy term. *Journal of Public Economics*. 118: 41-51.

Beshears, J., Choi, J., Laibson, D., Madrian, B., and Milkman, K. 2015. The effect of providing peer information on retirement savings decisions. *Journal of Finance* 70(3): 1161-1201.

Besley, T., and Case, A. 1995. Incumbent behavior: vote-seeking, tax-setting, and yardstick competition. *American Economic Review* 85: 25-45.

Bhargava, S. and Manoli, D. 2015. Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment. *American Economic Review* 105(11):1-42.

Brett, C., and Pinske, J. 2000. The determinants of municipal tax rates in British Columbia. *Canadian Journal of Economics*. 33(3): 695-714.

Borck, R., Fossen, F., Freier, R., and Martin, T. 2015. Race to the debt trap? – Spatial econometric evidence on debt in German municipalities. *Regional Science and Urban Economics* 53: 20-37.

Borck, R. Caliendo, M., and Steiner, V. 2007. Fiscal competition and the composition of public spending: Theory and evidence. *FinanzArchive* 63: 264-276.

Bordignon, M., Cerniglia, F., and Revelli, F. 2003. In search of yardstick competition: A spatial analysis of Italian municipality property tax setting. *Journal of Urban Economics*. 54: 199-217.

Brueckner, J.K., 1998. Welfare reform and interstate welfare competition: Theory and evidence. Assessing the New Federalism Occasional Paper No. 21, The Urban Institute, December.

Brueckner, J.K. 2003. Strategic interaction among governments: An overview of empirical studies. *International Regional Science Review*. 26: 175-188.

Brueckner, J.K., and Saavedra, L. 2001. Do local governments engage in strategic property-tax competition? *National Tax Journal*. 54(2): 203-203.

Bursztyn, L., Ederer, F., Ferman, B., and Yuchtman, N. 2014. Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica* 82(4): 1273-1301.

Burt, R. 1992. *Structural Hole.* Cambridge, MA: Harvard Business School Press.

Capello, R. 2009. Spatial spillovers and regional growth: A cognitive approach. *European Planning Studies* 17(5): 639-658.

Congressional Research Service. 2006. *Summary of the Pension Protection Act of 2006.* Washington, DC: Congressional Research Service.

Conley, J, and Dix, M. 1999. Optimal and equilibrium membership in clubs in the presence of spillovers. *Journal of Urban Economics* 46(2): 215-229.

Chetty, R., Friedman, J., Olsen, T., and Pistaferri., L. 2011. "Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records." *Quarterly Journal of Economics* 126(2):749-804.

Chetty, R., Friedman, J., and Saez, E. 2013. Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings. *The American Economic Review* 103(7): 2683-2721.

Chetty, R., Hendren, N., Kline, P., Saez, E. 2015. The economic impact of tax expenditures: Evidence from spatial variations across the U.S. SOI Working Paper. https://www.irs.gov/statistics/soi-tax-stats-soi-working-papers

Chetty, R., Looney, A., and Kroft, K. 2009. Salience and taxation: Theory and evidence. *The American Economic Review* 99(4): 1145-1177.

Chetty, R., and Saez, E. 2013. Teaching the tax code to EITC recipients. *American Economic Journal: Applied Economics*. 5(1): 1-31.

Currie, J. 2006. The take-up of social benefits. In *Public Policy and the Income Distribution,* edited by Alan. J. Auerbach, David Card, and John M. Quigley, 80-148. New York: Russel Sage Foundation.

DeSilva, S., Pham, A., and Smith, M. 2012. Racial and ethnic price differentials in a small urban housing market. *Housing Policy Debate*. 22(2): 241-269.

Duflo, E., Gale, W., Liebman, J., Orszag, P., and Saez, E. 2006. Saving incentives for low- and middle-income families: Evidence from a field experiment with H&R Block. *The Quarterly Journal of Economics*, 121(4), 1331-1346.

Duflo, E., Gale, W., Liebman, J., Orszag, P., and Saez, E. 2007. Savings incentives for low- and middle-income families in the United States: Why is the Saver's Credit not more effective? *Journal of the European Economic Association*, 5(2-3), 647-661.

Fajgelbaum, P., Morales, E., Serrato, J.C.S., and Zidar, O. 2019. State taxes and spatial misallocation. *Review of Economic Studies* 86: 333-376.

Finkelstein, A. 2009. E-Z Tax: Tax salience and tax rates. *Quarterly Journal of Economics* 969-1010.

Fogli, A., Guerrieri, V. 2019. The end of the American dream? Inequality and segregation in US cities. No. w26143. National Bureau of Economic Research.

Fortin, B., Lacroix, G., and Villeval, M.C. 2007. Tax evasion and social interaction. *Journal of Public Economics* 91(11-12): 2089-2112.

Funashima, Y., and Ohtsuka, Y. 2019. Spatial crowding-out and crowding-in effects of government spending on the private sector in Japan. *Regional Science and Urban Economics.* 75: 35-48.

Gale, W., Iwry, J. M., and Orszag, P. 2004. *Improving the Saver's Credit.* Brookings Institution.

Gale, W., Iwry, J.M, and Orszag, P. 2005. Making the Tax system Work for Low-Income Savers. *Tax Policy Issues and Options.* The Urban-Brookings Tax Policy Center, July.

Gallagher, K., and Muehlegger, E. 2011. Giving green to get green? Incentives and consumer adoption of hybrid vehicle technology. *Journal of Environmental Economics and Management* 61(1): 1-15.

Goldin, J. and Homonoff, T. 2013. Smoke gets in your eyes: cigarette tax salience and regressivity *American Economic Journal: Economic Policy* 5(1): 302-326.

Goldin, J., and Listokin, Y. 2014. Tax expenditure salience. *American Law and Economics Review* 16(1): 144-176.

Golgher, A., and Voss, P. 2016. How to interpret the coefficients of spatial models: spillovers, direct and indirect effects. *Spatial Demographics* 4:175-205.

Granovetter, M. 1973. The strength of weak ties. *American Journal of Sociology* 78(6):1360-1380.

Granovetter, M. 1985. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology* 91(3): 481-510.

Greene, W. 2003. *Econometric Analysis: Fifth Edition.* Prentice Hall, New Jersey.

Heim, B., and Lurie I.Z. 2014. Taxes, income, and retirement savings: Differences by permanent and transitory income. *Contemporary Economic Policy,* 32(3), 592-617.

Heyndels, B., and Vuchelen, J. 1998. Tax mimicking among Belgian municipalities. *National Tax Journal*. 51(1): 89-101.

Horowitz, J., and Manski, C. 1995. "Identification and robustness with contaminated and corrupted data." *Econometrica* 63(2): 281-302.

Hotz, V. J., and Scholz, J. K. 2003. The Earned Income Tax Credit. In *Means-Tested Transfer Programs in the United States,* edited by Robert Moffitt, 141-98. Chicago: University of Chicago Press.

Hoxby, C., and Bulman, G. 2015. How tax credits for higher education affect college-going. SOI Working paper; https://www.irs.gov/statistics/soi-tax-stats-soi-working-papers

Imbens, G., Kalyanaraman, K. 2009. Optimal bandwidth choice for the regression discontinuity estimator No. w14726. National Bureau of Economic Research.

Internal Revenue Service. 2019. Research, Applied Analytics & Statistics. Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2011-2013. Publication 1415. Washington, D.C.: Government Printing Press.

Jeszeck, C. 2015. *Retirement Security: Most Households Approaching Retirement Have Little Savings.* Washington, DC: General Accountability Office.

Kleven, H., and Waseem, M. 2013. Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan. *Quarterly Journal of Economics*, 128:669-723.

Koenig, G., and Harvey, R. 2005. Utilization of the Saver's Credit: An analysis of the first year. *National Tax Journal*, 787-806.

Larrimore, J. Mortenson, J., Splinter, D. 2017. Whose Child Is This? Shifting of Dependents Among EITC Claimants Within the Same Household. *National Tax Journal* 70(4): 737-758.

Le Gallo, J. 2014. Cross-Section Spatial Regression Models. Chapter from *Handbook of Regional Science*., eds. M.M. Fischer and P.N. Nijkamp. Springer-Verlag Berlin Heidelberg. pp 1511-1533.

Lee, David and David Card. 2008. "Regression discontinuity inference with specification error." *Journal of Econometrics* 142 (2): 655-674.

LeSage, J., and Pace, R.K., 2009. *Introduction to Spatial Econometrics*. Chapman and Hall/CRC Press. Boca Raton, FL.

McCrary, J. 2008. Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics*, 142(2):698-714.

Miller, B. and Mumford, K. 2015. The salience of complex tax changes: Evidence from the child and dependent care credit expansion. *National Tax Journal* 68(3): 477-510.

Ojede, A., Atems, B., and Yamarik, S. 2018. The direct and indirect (spillover) effects of productive government spending on state economic growth. *Growth and Change* 29(1): 122-141.

Owens, A. 2019. Building inequality: Housing segregation and income segregation. *Sociological Science* 6: 497-525.

Plueger, D. 2009. Earned income tax credit participation rate for tax year 2005. *IRS Research Bulletin*, 151-195, Washington, DC, Internal Revenue Service.

Polanya, K. 1968. *The Economy as Institutional Process.* In Economic Anthropology, E. LeClair, H. Schneider (eds). New York: Holt, Rinehart, and Winston.

Ramnath, S. 2013. Taxpayers' responses to tax-based incentives for retirement savings: Evidence from the Saver's Credit notch. *Journal of Public Economics,* 101, 77-93.

Reardon, S., Bischoff, K., Owens, A., Towsend, J.B. 2018. Has income segregation really increased? Bias and bias correction in sample-based segregation estimates. *Demography* 55.6: 2129-2160.

Revelli, F. 2001. Spatial patterns in local taxation: Tax mimicking or error mimicking? *Applied Economics* 33: 1101-1107.

Revelli, F. 2005. On spatial public finance empirics. *International Tax and Public Finance*. 12: 475-492.

Rork, J. 2003. Coveting thy neighbors' taxation. *National Tax Journal.* 56(4): 775-787.

Saad, L. 2015. Americans' money worries unchanged from 2014. Gallup news, published April 20, 2015. https://news.gallup.com/poll/182768/americans-money-worries-unchanged-2014.aspx

Saavedra, L. 2000. A model of welfare competition with evidence from AFDC. *Journal of Urban Economics*. 47(2): 248-279.

Saez, E. 2010. Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy* 2(3): 180-212.

Sallee, J. 2011. "The surprising incidence of tax credits for the Toyota Prius." *American Economic Journal: Economic Policy* 3(2): 189-219.

Scholz, J. K. 1994. The earned income tax credit: Participation, compliance, and antipoverty effectiveness. *National Tax Journal* 47(1): 64-87.

Sinai, T., and Gyourko, J. 2004. The (un)changing geographic distribution of housing tax benefits: 1980-2000. National Bureau of Economic Research Working Paper No. 10322.

Slemrod, J. 2007. Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives* 21: 25-48.

Solè-Ollè, A. 2006. Expenditure spillovers and fiscal interactions: Empirical evidence from local governments in Spain. *Journal of Urban Economics*. 59: 32-53.

Taubinksy, D., and Rees-Jones, A. 2018. Attention variation and welfare: Theory and evidence from a tax salience experiment. *The Review of Economic Studies* 85(4): 2462-2496.

Tobler, W., 1970. A Computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(Supplement), 234-240.

U.S. Census Bureau. TIGER/Line Shapefiles (machine-readable data files).

U.S. Department of Transportation. June 2011. Summary of Travel Trends; 2009 National Household Travel Survey. No. FHWA-PL-11-022. Washington DC.

Weber, M., and Bryant, V. 2005. The 1999 individual income tax return edited panel. In *2005 Proceedings of the American Statistical Association* 1, 59-166.

CHAPTER 9 BIOGRAPHY

Victoria L. Bryant graduated from Lake Braddock Secondary School, Burke, Virginia in 2000. She received her Bachelor of Arts from Virginia Tech in 2003. She received her Master of Arts from George mason University in 2008. She is employed as a Senior Economist at the Internal Revenue Service where she has worked since her graduation from Virginia Tech, barring a small stint as a local coffeeshop owner from 2005 through 2007. She has published and presented numerous papers covering a wide range of topics including taxpayer response to tax expenditures, statistical modeling, intergenerational income mobility, and retirement savings accumulation and distribution.