Socio-economic Diversity, Social Capital, and Tax Filing Compliance in the United States*

James Alm[a]
Jeremy Clark[b]
Kara Leibel[c]

Abstract: In this paper we present a rare empirical study of the determinants of tax filing compliance in the United States. We use county level data for the tax year 2000 and a panel of county and state level data for the tax years 2000 to 2006. We include explanatory variables identified in the rational compliance framework, including an enforcement index against identified non-filers, the audit rate of filers, and the average penalty rate for both filers and non-filers. We also examine the role of social capital on tax compliance. In particular, we test whether heterogeneity in household income, language, race, or religion can explain variation in filing rates. We find that non-filing rates tend to fall in the enforcement index in 2000 cross section analysis, but instead rise in the audit rate of filers in panel analysis. Non-filing rates also fall in the share of a county's population that is married or residentially stable, and rise in the share of county income from self-employment or public assistance and in the share of owner-occupied housing. Regarding social capital, non-filing looks to be increasing in heterogeneity by race, though not income or language. Non-filing may also be decreasing in heterogeneity by religious membership, though we have only cross-section evidence.

Keywords: tax filing compliance, social capital, heterogeneity

JEL classification codes: H24; H26; H31

[a] Department of Economics, Tulane University, 6823 St. Charles Avenue, 208 Tilton Hall, New Orleans, LA, 70118. Phone: +1 504 862 8344. Email: jalm@tulane.edu.

[b] Department of Economics and Finance, University of Canterbury, Private Bag 4800, Christchurch, 8020, New Zealand. Phone: 011 643 358 5667. Email: jeremy.clark@canterbury.ac.nz. Author for correspondence.

[c] Office of Research, Internal Revenue Service, 1111 Constitution Avenue, N.W., 3rd Floor/017, Washington, DC 20224. Email: kara.e.leibel@irs.gov.

1. Introduction

The last decade has seen a rise in studies identifying a negative correlation, and possibly causation, between measures of social heterogeneity and various indicators of "social capital". Social capital is defined by Putnam (2007) as people's beliefs and actions that contribute to "…social networks and the associated norms of reciprocity and trustworthiness". The dimensions of heterogeneity studied for their effect on social capital have included race or ethnicity, income, education, birthplace, first language, and religion. Studies have measured social capital using indicators such as people's propensity to volunteer, vote, join organizations, contribute to fundraisers, trust others, or support welfare programs (Alesina and La Ferrara 2000, 2002; Luttmer 2001; Costa and Kahn 2003a, 2003b; Miguel and Gugerty 2005; Leigh 2006; Putnam 2007; Gustavsson and Jordahl 2008; Clark and Kim 2012). Other studies have found government responses to heterogeneity that are consistent with a negative relationship between social heterogeneity and social capital. For example, Poterba (1997) finds that the negative relationship between U.S. state-level per-child education spending and the proportion of elderly in the state is exacerbated when the elderly are predominantly from a different racial group than the school-aged population. Alesina, Baqir, and Easterly (1999) find that increased racial heterogeneity reduces local government provision of core public goods such as roads, sewerage, and education. This paper examines the impact of social heterogeneity (and other factors) on a largely unexamined issue: tax compliance.

We make two principle contributions. First, we examine the effect of social heterogeneity on a compulsory rather than a voluntary indicator of social capital, tax filing compliance behavior. Second, we provide a rare empirical study on the determinants of tax filing compliance, using data derived from actual individual filing decisions.

Perhaps for reasons of data availability, most studies that have looked at heterogeneity's effects on social capital have used indicators of people's *voluntary* beliefs and actions. Few studies have tested for links between heterogeneity and indicators of people's *compulsory* actions. A rare example is Vigdor (2004), who finds that households' census form return rates for the 2000 United States census were lower in census tracts that were more heterogeneous by race, age, or education.

The effect of heterogeneity on tax compliance has been examined in a much more limited fashion. Three studies have used aggregate data on indirect measures of compliance, making cross-national comparisons of perceived tax compliance or attitudes toward tax compliance based on survey data. La Porta et al. (1999) and Alesina et al. (2003) both use a national index of tax compliance constructed from the 1996 Global Competitiveness Report, and each study finds a negative correlation between ethnic heterogeneity and compliance across countries.[1] Li (2010) uses World Values Survey data on whether cheating on taxes is justifiable, and similarly finds more sympathy for non-compliance in countries with greater ethnic heterogeneity. While these studies are suggestive of an effect, we are not aware of any paper that has examined the effects of heterogeneity on *actual* tax compliance based on *individual* filing decisions, simply because such data are not widely available.

Thus, the first contribution of our study is to plug the hole in the heterogeneity and social capital literature in the area of actual tax compliance. We examine the co-variation of heterogeneity and tax filing compliance rates in the United States between 2000 and 2006. We measure compliance with the United States' federal tax-filing requirements for the individual

---

[1] The Global Competitiveness Report relies exclusively on survey data regarding tax evasion, using respondents' agreement with the statement "[t]ax evasion is minimal in your country."

income tax using Internal Revenue Service (IRS) data on identified non-filers.[2] In recent years

the IRS has identified approximately nine million cases annually of individuals who have either

stopped filing without clear cause, or for whom third-party income reports indicate filing should

be expected.[3] With access to these data, we are able to estimate known non-filing rates for

almost all 3,140 counties in the United States and can therefore examine the effect of

heterogeneity at either the county or the state level on non-filing, while controlling for other

factors shown in the tax literature to predict non-compliance (Dubin, Graetz, and Wilde 1990;

Beron, Tauchen, and Witte 1992; Erard and Ho 2001; Gentry and Kahn 2009). Based on data

availability, we examine the effect of heterogeneity by race at the county level for all 7 years, by

language and household income at the county level for 2000 and at the state level for all 7 years,

and by religion at the county level for 2000 only.

Heterogeneity aside, the second contribution of our study is to provide a rare

investigation of the determinants of compliance with tax filing requirements in the United States.

The revenue lost from non-filers is sizable. The IRS estimates that in 2006 non-filing accounted

for over 8 percent (or $25 billion) of the gross individual income tax gap of $296 billion (IRS

2012, p. 2).[4] The absence of compliance data has often limited researchers in their ability to test

---

[2] Our discussion of filing compliance is limited to the individual income tax, although a broader consideration of non-filing could include other taxes such as the estate tax, corporate tax, or employer payroll tax. The IRS defines individual income tax non-filers as individuals who were legally required to file an income tax return and for whom returns had not been filed by the due dates of the returns or extended due dates.

[3] There is a difference between the IRS definition of non-filers and the set of individuals that the IRS *identifies* as non-filers. The identified non-filers are those discovered by a largely automated procedure that matches current filing data with prior filing information and third-party-reported income information. The concept of non-filer that we use in this paper is the set of individuals identified by the IRS in this automated procedure. This will not be a complete measure of non-filers for two reasons: (1) the IRS may discover some non-filers through other methods, such as audits of related taxpayers, although this happens fairly infrequently relative to those identified through the automated procedure; and, more importantly, (2) there likely exist other United States residents who have never filed a tax return (when required by law to do so), and for whom either the IRS receives no third party (e.g. bank or employer) income reports or the IRS receives reports with minimal income reported despite higher actual income earned. These are the "unknown non-filers", who cannot be addressed by our study.

[4] The non-filing gross tax gap is an estimate of the tax unpaid by non-filers according to the IRS definition, which is broader than the one we use in this paper.

the various theoretical models of tax compliance. Even when these data have been available, they have typically been for *reporting* non-compliance, focusing on individuals who file tax returns but do not declare all taxable income (Clotfelter 1983; Beron, Tauchen, and Witte 1992; Dubin, Graetz, and Wilde 1990; Gentry and Kahn 2009). There are very few studies (empirical or theoretical) on the determinants of the logically prior form of *filing* non-compliance, or whether individuals file tax returns when required by law to do so. Again, lack of data on non-filing "ghosts" has largely been responsible for this omission. One notable exception is work by Erard and Ho (2001). Using data from the IRS Taxpayer Compliance Measurement Program (TCMP), Erard and Ho (2001) compare 2,195 located non-filers in 1988 whose delinquent returns were audited, with a sample of 13,500 filers' audited returns. Their results indicate that the estimated 7.9 million ghosts in the United States in 1988 had income that was on average only about half that of filers, and that 43 percent of all non-filers made at least some form of prepayment, covering on average about half their income tax liability. In total, Erard and Ho (2001) estimate that non-filers accounted for a modest $7.8 billion in unpaid income and self-employment taxes in 1988, representing a tax gap only 15 percent as large as that for filers who underreported income. The one other empirical study on non-filing of which we are aware, by Dubin, Graetz, and Wilde (1990), examines the effects of IRS audit rates on filing and reporting compliance using audit data from the Annual Report of the Commissioner of Internal Revenue, 1977-1986. They find that *reporting* compliance is positively associated with audit rates of filed returns, but that *filing* compliance is negatively associated (possibly because higher audit rates increase incentives to avoid being audited by keeping out of the system altogether). Other than audit rates, Erard and Ho (2001) and Dubin, Graetz, and Wilde (1990) find evidence that filing

5

rates are affected by self-employment, occupation, education, age, unemployment, and state-level tax burden. Neither study examines the effect (if any) of heterogeneity on filing rates.

The next section discusses our theoretical framework for examining the role of "rational compliance" variables and social heterogeneity on reporting and filing compliance. We then present our data and empirical estimation strategy in section 3. In section 4 we present our results, while Section 5 concludes.

2. Theoretical Framework

Most theoretical analysis of the individual's tax compliance decision focuses on his or her income reporting decision, assuming that the individual has already chosen to file a return. This approach begins with the economics-of-crime model of Becker (1968), first applied to tax compliance by Allingham and Sandmo (1972). Here a rational individual maximizes the expected utility of the tax evasion gamble, weighing the benefits of successful under-reporting against the risky prospect of detection and punishment. The individual pays taxes because he or she is afraid of getting caught and penalized if some income goes unreported. This "portfolio" approach gives the plausible result that compliance depends upon audit rates and fine rates. Indeed, the central point of this approach is that an individual pays taxes because – and *only* because – of this fear of detection and punishment.[5]

More formally, an individual who has already decided to file a tax return must choose the amount of income to report, *R*, so as to maximize expected utility:

(1)   Reporting Decision: *EU = (1-p) U(I-tR-C) + p U(I-tR-(1+f)t(I-R)-C).*

---

[5] See Cowell (1990), Andreoni, Erard, and Feinstein (1998), Alm (1999), Slemrod and Yitzhaki (2002), Sandmo (2005), and Alm (2012) for comprehensive surveys and discussions of this literature.

Here E is the expectation operator, $U(\ )$ is the utility function, p is the probability that a tax return is selected for audit, I is the individual's "true" income, t is the tax rate on reported income, f is the penalty rate on undeclared taxes, and C is the financial (and time) burden of preparing and filing a tax return. The expected utility in equation (1) is the sum of two terms: the first if the individual files a tax return at cost C, reports income R, pays taxes tR, and is not audited with probability $(1-p)$; the second if the individual is caught with probability p and is forced to pay all taxes on unreported income as well as a proportional fine f.[6] Optimization with respect to R proceeds using standard methods, and comparative statics results are easily derived.

However, the individual's *filing* decision is also of interest, and has seldom been examined. To address the filing decision, the individual must compare the expected utility from filing with optimal income reporting to the expected utility from non-filing. The risk of enforcement for non-filing is likely to be significantly different from the risk of audit for a return that is "in the system"; similarly, the penalty for failing to file a return is likely to be different (and higher) than a penalty imposed on detected underreporting.

To address both filing and reporting decisions, we retain but modify the reporting framework developed in Erard and Ho (2001), Alm et al. (2010), and Kleven et al. (2011). An individual who decides to file a tax return and report income R has expected utility equal to equation (1). An individual who chooses not to file a tax return (and so who reports no income) has expected utility equal to:

(2)   Filing Decision: $EU = (1-p^*)\,U(I) + p^*\,U(I-(1+f^*)tI)$.

---

[6] In this framework, we ignore the distinction between the probability of audit and the probability of detection of underreported income; here we assume these are identical.

Here p^* is the probability that an individual with income I who has not filed a return is apprehended by enforcement activity, and f^* is the penalty imposed on detected non-filers. Note that p^* is likely to be different than p, that f^* is likely to exceed f, and that an individual who does not file a return escapes the filing cost C. The individual will thus compare the expected utility from filing and reporting the optimal R in equation (1), with expected utility from not filing in equation (2), and choose the greater.

This framework can be easily expanded to incorporate other relevant considerations. We focus on three.

The first is the possibility that the probability of detection for filers depends upon the amount of reported income, or $p(R)$. We assume that $p'(R) < 0$, so that the probability of audit decreases with reported income. This assumption follows from standard audit procedures in tax administrations around the world.

The second and related consideration is third-party reporting and its effects on the probabilities of audit of filers or enforcement against non-filers. A standard feature of individual income tax systems is that a third-party (e.g., the individual's employer or financial institution) reports the relevant part of an individual's taxable income to the tax authority (and often also withholds income taxes on this reported taxable income). This information increases significantly the chances that an individual who either does not file a return at all or who underreports income on a filed return will be detected. To address this, we can partition true income I between what is subject to third-party reporting, I_{TPR}, and what must be self-reported, I_{SR}. The audit rate for returns p is now assumed to be a function both of reported income R and of income subject to third-party reporting I_{TPR}. More precisely, $p(R, I_{TPR})$ is assumed to be a non-linear function whose value depends upon the relative magnitudes of R and I_{TPR}. If $R<I_{TPR}$,

then the audit probability will be very high because the tax authority knows from third-party information that the individual is under-reporting income; if $R \geq I_{TPR}$ then p will be much lower and decreasing rapidly. The presence of third-party information also affects the probability of enforcement against non-filers, or p^*, now written in the form $p^*(I_{TPR})$. If the tax authority has any information that an individual has some income I_{TPR} that is greater than the filing requirement but has not filed a tax return, then p^* will be very high; in the absence of this information, the probability will remain low and likely be close to 0.

A third consideration is the role social norms play in individuals' reporting and filing decisions. The portfolio model (even with the above extensions) focuses entirely on the financial considerations of reporting and filing. However, there is much evidence that individuals are influenced by the social context in which they make decisions. For example, there appears to be what may be termed a "social norm" of tax compliance (Elster, 1989). Although difficult to define precisely, a social norm can be distinguished by the feature that it is process-oriented, unlike the outcome-orientation of individual rationality. A social norm therefore represents a pattern of behavior that is judged in a similar way by others and is therefore sustained in part by social approval or disapproval. Consequently, if others behave according to some socially accepted mode of behavior, then the individual will behave appropriately; if others do not so behave, then the individual will respond in kind. The presence of a social norm is consistent with a range of conceptual frameworks, whether they rely upon fairness, trust, social customs, tax morale, appeals to patriotism or conscience, or feelings of altruism, morality, guilt, and alienation. Incorporating social norms suggests that, all else equal, an individual will comply as long as he or she believes that compliance is the social norm (however defined); conversely, if

non-compliance becomes pervasive, then the social norm of compliance disappears.[7] To the extent that heterogeneity affects social capital in the form of the frequency of people's interactions with others, it may affect either the social norms regarding compliance, or the weight individuals attach to complying with those norms.

The simplest way of introducing social norms is to assume that there is a non-pecuniary (or psychic) cost associated with evading one's own tax liability if one is not caught, as captured by the variable γ.[8] We posit that the greater is a society's social capital, the greater is the psychic cost to an individual of cheating on his or her tax liability, or $\gamma(Social\ Capital)$ with $\gamma'() > 0$. To the extent that heterogeneity affects social capital, it would then affect γ, either by changing social norms directly or by changing the importance individuals assign to abiding by social norms. Note that the psychic cost associated with cheating arises only if one is cheating and is not caught cheating, as assumed and analyzed by Gneezy and Rustichini (2000). Thus, a taxpayer who complies fully will experience no change in utility from the psychic cost of evasion.

Putting these three considerations together, equations (1) and (2) are modified as follows. The reporting decision becomes:

(1)'   Reporting Decision: $EU = (1-p(R, I_{TPR}))\ U(I-tR-C-\gamma) + p(R, I_{TPR})\ U(I-tR-(1+f)t(I-R)-C)$,

and similarly the filing decision becomes:

(2)'   Filing Decision: $EU = (1- p^*(I_{TPR}))\ U(I-\gamma) + p^*(I_{TPR})\ U(I-(1+f^*)tI)$.

---

[7] See Cowell and Gordon (1988), Gordon (1989), Myles and Naylor (1996), Kim (2003), Fortin, Lacroix, and Villeval (2007), and Traxler (2010) for examples of social interactions theory as applied to tax evasion.

[8] We assume that there is no psychic benefit from paying the tax. This simplifies the analysis, and may be justified if one takes the "paying taxes is a duty" viewpoint. It is straightforward to introduce such a psychic benefit.

As before, the individual chooses whether or not to file a tax return by comparing the value of expected utility from (2)' with that from (1)' with R optimized.

In the face of these many elements, the impacts on the individual's compliance decisions of variables like the tax rate and income reflect complicated interactions of income and substitution effects, and unambiguous comparative statics results often become elusive. For the enforcement variables (e.g., the probabilities of audit of filers or enforcement against non-filers and the penalty rates), the comparative statics results are more straightforward, indicating that an increase in, say, the enforcement probability against non-filing decreases the payoff to non-filing. The impact of heterogeneity via social norms is not clear-cut, because it would change the psychic cost of cheating in both the reporting and filing decisions.[9]

Regardless, however, it is straightforward to demonstrate in this framework that the likelihood of non-filing will depend upon an array of fiscal factors, like the probabilities of audit for filers, enforcement for non-filers, the penalty rates for filers and for non-filers, the tax rate, and income. The composition of income between third party-verified and unverified sources will affect both the probability of audit and of detection and enforcement against non-filing. Non-filing will also depend upon unobserved preferences or costs associated with filing that may be proxied by socio-demographic variables (e.g., marital status, education, first language, race, or age). Finally, heterogeneity may affect social capital and thus social norms regarding compliance. Heterogeneity can be captured by dispersion measures of salient characteristics that people notice in those around them, such as income, first language, race, or religion. The next section presents our empirical framework for estimating these various impacts.

---

[9] More precisely, recall that an individual compares expected utility in the two cases (or equation (1)' versus equation (2)'). An increase in the psychic cost lowers expected utility in both bases, by an amount that depends upon the product of the probability of non-audit and the marginal utility of income in the relevant state. In general, it is not possible to determine whether this change in expected utility is greater in one case versus the other.

3. Data and Estimation Strategies

3.1. Data

To address the determinants of filing compliance empirically, we have assembled data on IRS-identified non-filing rates for counties in the United States for 2000-2006, as well as IRS-sourced measures of "rational compliance" variables such as audit rates of filed returns, likelihood of enforcement initiated on identified non-filed returns, and effective penalty rates on delinquent filers and on filed returns containing misreported income or other errors. To these we have added census and other non-IRS variables used in the tax compliance literature to measure social heterogeneity, income levels, source of income as it affects third-party reporting and the probability of detection, and demographic characteristics that signal potential pecuniary benefits from filing (e.g. the Earned Income Tax Credit (EITC)) or that proxy for unobserved preferences regarding compliance. All variables are described in Appendix A, and summary statistics are provided in Tables 1 and 2. Here we summarize the key variables.

[Tables 1 and 2 about here.]

To generate an (identified) non-filing rate, we take the total number of individuals identified by the IRS as non-filers for a given tax year in a county, and divide it by the total number of individual filers plus non-filers in that county.[10] We construct a "broad" measure of non-filing that includes late-filers, as well as two "narrow" measures that are restricted to those who do not file and are thought to have a tax liability either greater than $0, or greater than $500. Our broad measure thus attempts to capture *all* (identifiable) non-filers at the cost of including

---

[10] Married jointly filed returns are counted as representing two individuals.

some who may not have needed to file. Our narrow measures attempt to avoid including the ex post "innocent", at the cost of excluding some of the ex post "guilty."

Our "rational compliance" measures are constructed as follows. For p, the probability of enforcement that is relevant to the income reporting decision, we follow standard practice in the literature and use the probability of being audited, or an "audit rate". We define this rate as the number of individual income tax audits initiated in calendar year t in a given county divided by the number of returns from that county filed in year t (reflecting income earned in calendar/tax year t-1).[11] We assume that this year t audit rate affects the filing decision as of the filing deadline of April 15th, t+1 and therefore the rate of non-filers for tax year t.[12]

Our model dictates that we also need a measure of p^*: the probability of enforcement against non-filers. Yet there is no obvious precedent for this in published research, nor is there an appropriate measure reported by IRS, to our knowledge.[13] Unlike filed returns, non-filers are not technically audited, so we cannot use an audit rate for this; instead we need a measure that

---

[11] Our constructed audit rate differs from both the IRS' standard definition of the audit rate (see below) and from what might seem the theoretical ideal: the percent of returns filed for a given tax year that end up being audited. Although the logic of the latter measure is appealing, it is not practical here. Audits can be initiated up to several years after the filing of a return, so that the "true" audit rate for returns filed in year t would only be known several years later. It thus could not be expected to influence decisions made in year t.

The "true" audit rate is similarly impractical for IRS operational decision-making. Instead, the standard approach taken by the IRS (and thus typically used in other papers) is to define the audit rate as the number of individual income tax audits *closed* in *fiscal* year t divided by returns filed in calendar year *t-1* (reflecting income earned in calendar/tax year *t-2*).

Our measure instead counts audits by when they open, closely reflecting when the taxpayer first becomes aware of them. Our denominator captures tax filing behavior during the same time period as these initiated audits.

[12] The IRS does not consider taxpayers with valid extensions to file after April 15th to be non-filers if they file before the extended deadline. Conversely, taxpayers who do not request extensions and file after April 15th are considered non-filers. In the analysis that follows we use multiple definitions of non-filers that include or exclude the latter.

[13] Plumley (1996) uses a measure based on the IRS' issuance of targeted non-filer notices, and GAO (2003) provides an "enforcement contact rate" for several types of IRS enforcement: non-filer as well as audit, math error, and document matching. However, neither of these measures seems appropriate for our purposes in that we feel a "notice" or "contact" is not sufficient to be considered enforcement in the context of non-filing. The GAO contact rate also uses the number of filed returns as the denominator, which again is not appropriate for our analysis.

better reflects the enforcement process as it applies to identified non-filers.[14] We have chosen to consider "enforcement" to be the point at which the IRS makes an assessment of unpaid taxes, penalties, and interest against the non-filer.[15] The overall "enforcement index" is then defined as the number of non-filers (technically "delinquent returns") that have an initial financial assessment made against their account in fiscal year t divided by the number of (broad or narrow) individuals who were identified as not filing a return in calendar year t for income earned in the previous tax year.[16] As with our audit rate, we assume that this non-filer enforcement index will influence the filing decision made as of April 15th, $t+1$ and will be captured in the non-filer rate for tax year t.

The rational compliance model also indicates that the severity of the punishment, captured by proportional fines f and $f*$, should influence reporting and filing decisions.

---

[14] IRS enforcement for identified non-filers is typically conducted through correspondence with the taxpayer rather than an audit. Once the IRS identifies non-filers, which may be a year or more after the filing deadline, it will send notices requesting them to file. If the individual complies with the notice and files a return, the IRS will make an assessment of unpaid taxes, and most likely penalties and interest, when the return is processed. For those who do not comply with that request voluntarily, the IRS may prepare a Substitute For Return (SFR), which is a return prepared on the individual's behalf. When this SFR is "filed", the IRS will assess unpaid taxes, penalties and interest. Depending on the severity of the noncompliance, the amounts owed, and subsequent behavior by the individual, the IRS may also take steps to recover revenue through such actions as liens or seizure of property, and in rare cases, criminal prosecution.

[15] While it would be possible to choose a different point in the enforcement process, this one seems most consistent with our theoretical model. Up until the assessment is made, regardless of how many notices may have been sent by the IRS, the individual is still in the $(1-p*)$ state, with no additional cost having been incurred by non-filing. Once the assessment is made, the individual switches to the $p*$ state and has to pay the associated penalty and unpaid tax (or face more serious consequences). While it would be possible to use a later point in the enforcement process – say, the placing of a lien or seizure of property – we think this would be too narrow a measure as it would exclude many non-filers who were subject to some form of punishment through the penalties and interest that would have been assessed in addition to the unpaid tax.

It may be worth noting that the IRS will not necessarily pursue or ultimately make assessments against all identified non-filers. As with all IRS enforcement efforts, competing priorities, fixed resources, and the perceived egregiousness of the non-filing behavior will all contribute to how many and which non-filers have an assessment made against them.

[16] The numerator of our enforcement index comes from the IRS Enforcement Revenue Information System (ERIS), which tracks (on a fiscal year basis) cases where the IRS undertakes enforcement activity against a taxpayer to recover owed taxes, penalties and interest. ERIS assigns cases to different categories based primarily on the type of enforcement activity. The category that contains non-filers is referred to as "delinquent returns", which unfortunately includes a somewhat larger set of returns than our identified non-filers. For this reason, as well as the timing issues that also characterize our audit rate, our enforcement measure is not a true rate and has the potential to exceed 1. Accordingly, we consider our measure an "index" rather than a rate.

Regarding penalties for non-filing, the IRS imposes a failure-to-file penalty that is typically 5 percent of unpaid taxes for each month the return is late, up to a maximum of 25 percent. Interest may also be charged on any unpaid tax from the due date of the return.[17] Despite the stability of the penalty formula over time and counties, there may still be variation in effective penalty rates due to abatements, time to detection (and thus interest charged), and similar considerations. We thus define a "penalty rate for non-filers" by first taking all delinquent returns whose most recent enforcement revenue activity was in fiscal year t, and by then calculating the ratio of net penalties plus interest divided by net tax assessed for each individual. We average these to get a county-level penalty rate for non-filers. We construct a penalty rate for filers in an analogous manner.[18] Both the non-filer and filer penalty rates for year t are assumed to affect the filing decisions made as of April 15th, $t+1$.

It is important to note that all IRS-generated variables are censored for counties with fewer than 10 observations. For example, for the 3140 counties in tax year 2000, filing rates are available for 96 percent of the counties; non-filing enforcement indices are available for 93 percent, audit rates for 84 percent, and both types of penalty rates for 100 percent.[19] The counties absent from our analysis due to censoring are generally those sparsest in population, but may also reflect some that are most tax compliant.

In order to examine the role of heterogeneity in tax filing compliance, we create relevant measures for language, race, religion, and income. For qualitative language, race, and religion variables, we follow the social capital literature in using fragmentation measures, defined as:

---

[17] The interest rate charged is the federal short-term rate plus 3 percent, compounded daily.

[18] The data for both filer and non-filer penalty rates come from IRS Enforcement Revenue Information System, which means these rates are calculated only using those cases where there was some enforcement activity.

[19] Audit rates for filers are available for a slightly higher share of counties in subsequent years. For the 2000-2006 panel overall, 86 percent of county-years are available.

$$(3) \quad F = 1 - \sum_{i=1}^{n} s_i^2 .$$

A fragmentation index F begins by dividing a population into n categories. Each category's share of the population is squared, summed, and subtracted from 1. (Fragmentation can be thought of as 1 minus the Herfindahl Index of concentration.) F is bounded between zero (complete homogeneity) and an upper bound whose limit is 1 (complete heterogeneity); the exact upper bound is increasing in n. For language spoken at home, we define $F_{Language}$ over 3 categories: English, Spanish, and Other. For race, we define F_{Race} over the 7 categories used in the 2000 census: White Non-Hispanic, White Hispanic, Black Non-Hispanic, Black Hispanic, American Indian/Alaskan Native, Asian, and Hawaiian/Pacific Islander. For religion, we define $F_{Religion}$ using 6 aggregate categories of enumerated adherents at churches and other religious organizations available from the Association of Religion Data Archives (ARDA): Mainline Protestant, Evangelical Protestant, Roman Catholic, Orthodox, Other Religions, and Unclaimed. Finally, for quantitative income variables, we measure the dispersion of household income F_{Income}, using the Gini coefficient as provided by the Census and the ACS, which like the fragmentation index is bounded between 0 and 1 and increases in heterogeneity.

Note that of our four measures of heterogeneity, only the measure pertaining to race is available at the county level for all years. Language and household income measures are available at the county level for 2000 only, but at the state level for 2000-2006. Religion is available only for 2000, at the county level.

Other control variables thought to influence the filing choice are taken from the United States Census of 2000 at the county level, the American Community Survey (ACS) at the state level for 2000 to 2006, the Bureau of Economic Analysis at the state level, and the Area Resource File at the county level for 2000 to 2006. To proxy for eligibility for the EITC, which

provides tax credits for low-income working households according to household size, we have included the share of the population in the civilian labor force, the unemployment rate, average household size, median household income, the interaction of the household size and income, and this interaction squared. Demographic variables thought to proxy for unobserved preferences or costs of filing, such as age, sex, marital status, education, language spoken at home, stability of residence, home ownership rates, birthplace, and race, are similarly sourced from the Census or the ACS. Religious adherence rates at the county level are included for 2000 only, sourced from the ARDA.[20] Median (owner-occupied) home values are included to proxy for differences in cost of living that affect the real cost of income tax. Finally, following Dubin, Graetz, and Wilde (1990), we include average state individual income tax rates. Given the information sharing that exists between state and federal tax agencies, individuals in higher tax states may face higher incentives not to file federal returns.

3.2. Empirical Estimation Strategy

Cross section regressions based on tax year 2000 data have the advantage that almost all variables are available at the county level. Fixed effects regressions based on tax years 2000 to 2006 have the advantage of being able to control for unobserved but stable county characteristics, although some variables are available only at the state level. We thus proceed with both approaches. Since our dependent variables are non-filing rates in the 0-1 interval, we use a logit transformation to map them to the real line (Baum 2008).

Beginning with 2000 cross section estimation, we use the form:

(4) $\ln(y_{ij,2001}/1-y_{ij,2001}) = X'_{ij,2000}\beta + u_{ij,2001}$,

---

[20]   ARDA primarily reports adherents enumerated by religious organizations themselves, where adherents are defined to include full members, their children, and others who regularly attend services. As a result, individuals who self-identify with a religion but do not associate with a specific church or group are classified as "unclaimed" along with those having no religion.

where $y_{ij,2001}$ refers to the (broad or narrow) non-filing rate in county i of state j for tax year 2000 on or by April 15th 2001, $X_{ij,2000}$ is a vector of county level rational compliance variables, demographics, and state dummies for 2000, and $u_{ij,2001}$ is a random error. With limited correlations between our four heterogeneity measures (none exceeding |0.54|), we follow this with a specification that adds all of them simultaneously, or:

$$(4)' \qquad \ln(y_{ij,2001}/1 - y_{ij,2001}) = X'_{ij,2000}\beta + F'_{ij,2000}\delta + u_{ij,2001},$$

where $F_{ij,2000}$ is the vector of F_{Income}, $F_{Language}$, F_{Race}, and $F_{Religion}$ for county i in state j in 2000. With or without heterogeneity measures, we also include "sparse" and "full" specifications that vary the number of non-IRS control variables included. To aid with comparison across approaches, the criterion we use for excluding variables in "sparse" models is not an information criterion, but whether the variable is unavailable at the county level in subsequent panel analysis.

As is common in the empirical tax literature, we recognize the possibility that IRS audit efforts against filers and enforcement efforts against identified non-filers may be endogenously caused by non-filing rates in (4) and (4)'. Here, it is not possible for non-filing rates as of April 15th, 2001 to affect audit or enforcement activity in 2000. However, if the idiosyncratic error terms in (4) or (4)' are serially correlated over time, it can easily be shown that audit or enforcement measures in 2000 could be correlated with u_{2001} in those equations. We address potential endogeneity using five county-level instruments sourced from the IRS. These instruments need to be correlated with audit or enforcement activity initiated in 2000 (on tax year 1999 returns), yet collectively uncorrelated with $u_{2001.}$ Because the IRS commonly initiates audit or enforcement activity one or two years after a tax year, most of these instruments are averaged over 1999 and 2000.

The instruments we use are: (1) the audit rate of the nearest neighbor county, defined as the county whose population center (or 'centroid') is closest to a given county's population center; (2) the enforcement index of the nearest neighbor county, similarly defined; (3) the *number* of third-party information documents sent to an address in a county per filed return averaged over 1999 and 2000; (4) the number of *types* of third party information documents sent to an address in a county over the 23 types possible averaged over 1999 and 2000; and (5) the percent of filed returns claiming eligibility for the EITC for tax years 1998 and 1999. After testing that these instruments are sufficiently correlated with the audit and enforcement measures and sufficiently uncorrelated with the error term to be valid, we estimate generalized method of moments (IV-GMM) rather than OLS if endogeneity is indicated.

Moving to panel data estimation, we use fixed effects regressions that follow counties between tax years 2000 and 2006. Here some explanatory variables X are available at the county level ij, whereas others Z are available only at the state level j:

(5) $\ln(y_{ijt+1}/1 - y_{ijt+1}) = X'_{ijt}\beta + Z'_{jt}\theta + \alpha_{ij} + u_{ijt+1}$.

We also include year dummies. As before, with limited correlations between our heterogeneity measures (non exceeding |0.42|), we follow this with a specification that adds all available fragmentations simultaneously:

(5)' $\ln(y_{ijt+1}/1 - y_{ijt+1}) = X'_{ijt}\beta + Z'_{jt}\theta + F'_{Income,jt}\delta_1 + F'_{Language,jt}\delta_2 + F'_{Race,ijt}\delta_3 + \alpha_{ij} + u_{ijt+1}$.

Also as before, we include both "sparse" specifications that exclude Z, as well as "full" specifications that include Z with standard errors clustered to the state level. We address potential correlation between the audit rate of filers or the enforcement index against non-filers in year t with the error term in $t+1$ using three of the five instruments used in cross section: the two nearest neighbor rates and the percent of filed returns claiming EITC eligibility. As with

19

cross section estimation, after testing the validity of our instruments, we estimate and present IV-GMM results if endogeneity is indicated.

Lastly, we test whether the determinants of non-filing are relevant for more egregious non-filing. Specifically, we repeat the cross section and panel analysis using a stricter definition of non-filers: those initially estimated by the IRS to owe at least $500 for a given tax year.

4. Results

4.1. Simple Correlations

Descriptive statistics for all variables are provided in Tables 1 (cross section) and 2 (panel). We begin by examining simple correlations between non-filing rates and our four "rational compliance" measures. Figure 1 provides scatter plots of both broad and narrow (>$0) non-filing rates against the average penalty rate for pursued non-filers, the enforcement index against identified non-filers, the average penalty rate for detected non-filers, and the audit rate of filers, all for tax year 2000. As expected, both broad and narrow non-filing rates are negatively correlated with the enforcement index, with pairwise correlations of -0.23 and -0.26, respectively. They appear even more strongly correlated with the audit rate of filers, with pairwise correlations of 0.38 and 0.43, respectively. This might suggest, as Dubin, Graetz, and Wilde (1990) found, that a more severe audit regime for filers has the side effect of raising non-filing rates. In contrast, neither type of non-filing rate looks to be correlated with the penalty rates meted out to non-filers or filers. While not illustrated in Figure 1, these patterns mostly persist in the pooled data for tax years 2000 to 2006. Over all seven years, broad and narrow non-filing rates are moderately negatively correlated with the enforcement index (-0.23 and -0.26, respectively), more strongly positively correlated with the audit rate (0.39 and 0.44,

respectively), not correlated with the penalty rate for non-filers (0.02 and 0.00, respectively), and slightly positively correlated with the penalty rate for filers (0.08 and 0.07, respectively).

We consider next the simple correlations between non-filing rates and social heterogeneity. Figure 2 provides scatter plots of broad and narrow (>$0) county non-filing rates against household income inequality, and language, race, and religious fragmentation for tax year 2000. A moderate positive correlation appears between either definition of non-filing and income and language heterogeneity, ranging from 0.30 to 0.42. A stronger positive correlation appears between broad or narrow non-filing and racial fragmentation, at 0.68 and 0.71, respectively. Finally, a moderate negative correlation appears between non-filing and religious adherents fragmentation, at -0.23 and -0.24, respectively. When we move to the 2000-2006 pooled data, language and income heterogeneity measures are available only at the state level, which likely reduces their variation, and data on religion are not available. Here we find that the positive correlation between either measure of non-filing and income or language heterogeneity is reduced to range between 0.18 and 0.24. In contrast, the positive correlation between broad or narrow non-filing and (still county level) racial fragmentation remains high at 0.67 and 0.68, respectively.

To see whether these relationships between non-filing rates and IRS enforcement variables or social heterogeneity persist once other factors are controlled, we move next to regression analysis.

[Figure 1 about here.]

[Figure 2 about here.]

4.2. Cross Section Results

We begin with year 2000 cross section analysis. Recall that we estimate eight models: broad and narrow non-filing rates, without and with controls for fragmentation, and with sparse or full specifications. To address potential endogeneity in each specification between the audit rate for filers or the enforcement index for non-filers and the error term, we use the "ivreg2" commands in the software Stata 12.0. For all eight models we use two-step generalized method of moments (IV-GMM) to test for the joint validity of our five proposed instruments. This involves testing whether each instrument was individually significant in at least one first stage regression, whether all five together were sufficiently correlated with the audit rate and enforcement index to reject under-identification or weak identification, and whether all five were collectively uncorrelated with the error term, or over-identified. For all eight models each instrument is significant at the 5 percent level or better in at least one of the two first stage regressions, and under-identification could be rejected at the 1 percent level using the Kleibergen-Paap rank LM statistic. Weak identification posed a moderate problem for some models even with 5 instruments. In particular, Kleibergen-Paap rank F statistics could reject a 10 percent maximum IV relative bias for four models, but only a 20 percent bias for three, and only a 30 percent bias for the last (narrow non-filing without heterogeneity, full specification). Over-identification could not be rejected at the 10 percent level for all eight models using the Hansen J statistic. Thus our instruments are valid, albeit with some weakness.

Applying our instruments, exogeneity of the audit rate and enforcement index together could be rejected at the 5 percent level or better for all eight models. We therefore discuss our IV-GMM results, though our OLS results are reported in Appendix B.

Results are provided in Table 3. To aid in the interpretation of coefficients from this table, we see first from Table 1 that the "odds" of broad non-filing (or the probability of broad

non-filing .0392 divided by the probability of broad filing .9608) is .0408 for tax year 2000 at sample mean. From column (2) in Table 3, the coefficient on "Enforcement Index Non-filers" is -0.541, the exponential of which ($e^{-.541}$) is 0.582. This implies that counties with a maximum enforcement index value against identified non-filers in 2000 have an odds of broad non-filing that is 58 percent that of counties with a minimum enforcement index. For negative coefficients such as this, the inverse odds ratio 1.72 (or 1/0.582) may also be helpful, indicating that counties with minimum enforcement have odds of non-filing 1.72 times the odds of counties with maximum enforcement, or odds that are 72 percent higher.

Table 3 suggests that both broad and narrow non-filing rates are decreasing in IRS enforcement efforts, as expected, but are not significantly affected by other rational compliance variables. The lack of effect from the audit rate of filers is particularly surprising in light of the strong simple correlations presented earlier. This result could stem from endogeneity being addressed, or it could be the result of omitted variable bias that is almost certainly present in all cross section models. Among other controls, non-filing rates tend to be falling in median household income, share aged 15-19 or 65 or older, share in the labor force, household size, unemployment, share married, residential stability, and share associated with Evangelical Protestant churches. Non-filing rates tend to be rising in share aged 45-64, some racial shares, share in poverty, share with some college/associate degree, share of income from self-employment or public assistance, share foreign born, median home value, share of homes that are owner occupied, and the interaction of household income with household size.

[Table 3 about here.]

Finally, when we add our four heterogeneity measures (columns 1 versus 3, 2 versus 4, 5 versus7, and 6 versus 8), the estimated effects of our existing variables remain stable, though less so for the shares underlying the language, race, and religion fragmentation measures. The loss of effect for an underlying language, race, or religion share when its corresponding fragmentation measure is added could suggest that these factors are affecting non-filing rates via their effects on heterogeneity and social norms, rather than reflecting differing propensities to file per se.

With this as background, the effect of heterogeneity on non-filing rates in cross section is mixed. Racial fragmentation looks to increase non-filing, with a move from complete homogeneity to complete heterogeneity increasing the odds of broad non-filing in column (4) by 49 percent (or $e^{.400} = 1.49$), and the odds of narrow non-filing in column (8) by 53 percent. (The odds of narrow non-filing are .0238 at sample mean.) Household income inequality and language fragmentation look to have no effect, while religious fragmentation looks to decrease non-filing. In particular, a county with complete religious heterogeneity would have odds of broad (narrow) non-filing 84 percent (88 percent) that of a county with complete homogeneity.

While our cross section results are for the most part plausible, and rest on quality census and IRS data, it is well known that unobserved influences on non-filing may be causing misleading correlations between included variables and biased coefficients. To better control for unobserved, time-invariant, county-level influences on non-filing, we move next to fixed effects panel estimation.

4.3. Fixed Effects Panel Results

As with our cross section analysis, we estimate eight models that follow counties through the tax years 2000 to 2006: broad and narrow non-filing rates, without and with controls for fragmentation, and with and without variables available only at the state level. We again start by

addressing potential endogeneity in each model between either the audit rate for filers or the enforcement index for non-filers, and the error term. For each model, we again use IV-GMM to test for the joint validity of our three instruments described earlier. For all eight models each instrument is individually significant in at least one of the two first-stage regressions at the 1 percent level or better. Under-identification could always be rejected at the 1 percent level, and weakness in the correlations between the instruments and the audit and enforcement measures is much less of a concern than in cross section. In particular, the Kleibergen-Paap rank Wald F statistics could easily reject a 10 percent maximal IV size for all eight models. On the other hand, while over-identification could not be rejected at the 10 percent level for 6 of 8 models, it could not be rejected only at the 5 percent level for the remaining 2 models (narrow sparse, with or without heterogeneity). Thus for 6 of 8 models our instruments are sufficiently uncorrelated with the error term as to be valid, but in 2 cases there is some concern.

With this as background, exogeneity of the audit and enforcement measures together could be clearly rejected for 5 of 8 models at the 5 percent level or better, rejected for 2 of 8 models at the 10 percent level, and rejected for the last model (narrow full with heterogeneity) at a p value of 0.111. In what follows we therefore emphasize the IV-GMM results; the alternative specifications are reported in Appendix B.

[Table 4 about here.]

As Table 4 indicates, some of our findings change under fixed effects, while others remain the same. Before considering these, recall that the move to fixed effects has the advantage of controlling for unobserved county characteristics, but the disadvantage that some

control variables are now at the state level, with attendant loss in variation. In addition, some measured characteristics may simply not vary much over a seven year period, making it difficult for fixed effects to discern their effects.

Beginning with rational compliance variables, we find that the enforcement index against non-filers now has only a small negative effect on broad non-filing, and loses any significant effect on narrow non-filing. In contrast, the audit rate for filed returns now has a strong positive effect on non-filing rates, more in keeping with the simple correlations shown initially. From column (6), counties with a maximum audit rate of filers have an odds of narrow non-filing that is ($e^{2.57}=$) 13.07 times that of counties with a minimum audit rate, a striking 1207 percent increase! (The odds of narrow non-filing over the 7 years are .0208 at sample mean.) Among other significant control variables, non-filing tends to be decreasing in the share male, residential stability, and the share married. Non-filing tends to be increasing in the share unemployed, share of income from public assistance, and share of houses that are owner occupied, and share aged 65 or older. More subtly, *narrow* non-filing is increasing in the share of income from self-employment or other sources, but *broad* non-filing is not. Conversely, broad non-filing is decreasing in the share of income from Social Security and increasing in the average state income tax rate, but narrow non-filing is not. The effect of race is also subtle: non-filing rates look to be increasing in the share Black Non-Hispanic when racial fragmentation is not controlled for (models 2 and 6), but not to be affected when it is (models 4 and 8).[21] The same pattern is observed for share White Hispanic for broad non-filing, though narrow non-filing is increasing in this share even when fragmentation is controlled for (model 8). Since the share White Hispanic is alone among racial shares in explaining variation in narrow non-filing rates

---

[21] We focus here on full models, because race variables are available at the county level for all years, unlike variables for language, income inequality or religion.

once fragmentation is controlled for, this might suggest that the share is proxying for citizenship or legal immigration status.

In contrast to cross section, fixed effects analysis does not find that non-filing is consistently affected by median household income, share aged 45-64, share in the civilian labor force, household size, or share in poverty, despite the fact that these variables are measured at the county level. Fixed effects analysis also does not find that non-filing is affected by share with some college/associate degree or by real median home value.

Turning finally to the effect of heterogeneity, we see that non-filing is again increasing in racial fragmentation[22], and again is not affected by language fragmentation or real household income inequality. The magnitude of these effects is greater for broad non-filing (that includes late filers or those not owing money) than for narrow non-filing. In particular, a move from complete homogeneity to heterogeneity increases the odds of broad non-filing in column (4) by $(e^{.740} =)$ 2.10, or 110 percent, but the odds of narrow non-filing in column (8) by 1.67, or 67 percent.

4.4. Stricter Definitions of Non-filing

Given that even those classified as "narrow" non-filers may still be estimated to owe only a few dollars, we check finally if our results change when we further narrow our focus to non-filers thought to owe at least $500 for a given tax year. Using the exact same empirical approach as for our earlier two definitions of non-filers, we find that the same instruments are valid for cross section and fixed effects specifications.[23] With endogeneity indicated as before, we present

---

[22] With exogeneity rejected only at the .111 level, the non IV result for model (8) in Appendix B may also be relevant. It finds that narrow non-filing is not significantly increasing in racial fragmentation. The coefficient there on racial fragmentation is .350, but its p value is .129.

[23] The one problem for instrument validity concerns models (5) and (7) of Table 5, which are panel, sparse, without and with heterogeneity measures. Hansen J statistics indicate our three instruments are correlated with the error term. We could not find a combination of instruments that was not correlated. With endogeneity indicated at the 1

IV-GMM results in Table 5. In general, these results are similar for the stricter definition of narrow non-filing as for our original narrow definition, with one exception. Stricter narrow non-filing again decreases in the enforcement index in cross section, but instead increases in the audit rate in fixed effects. Strict narrow non-filing consistently increases in the share of income from self-employment, public assistance, and other sources, as well as in the share of housing that is owner-occupied, but it decreases in residential stability. Particularly in panel estimation, racial shares look to matter until fragmentation is controlled for, just as before, though the share Asian retains a negative effect. One important change in results with a stricter definition of narrow non-filing is that racial fragmentation loses statistical significance in panel estimation (model (8))[24], although it retains significance in cross section results (model (4)). Viewed against the panel results for our earlier two definitions of non-filing, we see a pattern emerge. Racial fragmentation is seen to have the largest magnitude of effect on broad non-filing rates (model (4) of Table 4), a smaller magnitude of effect on narrow non-filing rates (model (8) of Table 4), and finally an even smaller but now insignificant effect on stricter narrow non-filing rates.

[Table 5 about here.]

5. Discussion and Conclusion

This paper has attempted to make two contributions. First, it has joined Dubin, Graetz, and Wilde (1990) and Erard and Ho (2001) in providing an empirical study of the determinants of tax filing compliance in the United States. Using actual individual United States tax data aggregated to the county level for the tax years 2000 to 2006, we have found that tax non-filing

percent level, we still present IV-GMM results for models (5) and (7) in Table 5. Corresponding non-IV results were generated, and did not differ substantively from what we discuss below.
[24] Though racial fragmentation retains significance in the sparse panel specification, there are reasons to doubt this result. The instruments used to generate the IV-GMM results for model (7) are invalid (see note 16), and racial fragmentation is not significant in the analogous non-IV regression.

rates have varied in ways that are consistent with augmented versions of the rational compliance model, though with some variation between cross-section and fixed effects results. In particular, with endogeneity of audit and enforcement measures addressed using instrumental variables, non-filing rates look to decrease in an enforcement index against identified non-filers in tax year 2000 cross-section, and are not affected by the audit rate of filed returns. In contrast, non-filing rates only slightly decrease in enforcement effort in fixed effects panel estimation, and only for the broad definition of non-filers. Rather, in fixed effects estimation, non-filing rates are sharply increasing in the audit rate of filed returns.

Among other control variables, some have effects on non-filing rates that are consistent across cross-section and fixed effects. Non-filing rates tend to be increasing in the share of a county's income that comes from self-employment or public assistance, and in the share of housing that is owner-occupied. They also tend to be decreasing in residential stability and in the share married. Other variables, however, have estimated effects that differ between cross-section and fixed effects. Covariates that proxy for EITC eligibility, for example, such as share in the labor force, household size, unemployment, and the interaction between household size and income, appear to vary with non-filing rates in predictable ways in cross-section regressions, but have little or no effect in panel regressions. Similarly, non-filing looks in cross-section to be decreasing in median household income, and increasing in the shares of people aged 45-64, with some college/associate degree, in poverty, or foreign born. However, none of these effects are found using fixed effects panel estimation. Finally, some variables matter for some definitions of non-filing, but not others. In particular, the share of county income from the census' "other sources" category does not explain variation in broad non-filing rates, but positively predicts narrow non-filing rates, more-so for those estimated to owe at least $500.

Why do our cross-section and panel findings differ? Conceptually, cross-section regressions based on tax year 2000 differ from fixed effects in having all variables available at county level, primarily from population counts. However, being based on a single year, effects are estimated only from variation *between* counties, and cannot control for counties' unobserved, stable characteristics that may be influencing compliance. Fixed effects for tax years 2000-2006 can effectively control for omitted county-level characteristics, but must use some (non-IRS) control variables defined at the state level and/or based on sampling estimates. Being based on variation *within* counties, fixed effects may also have trouble identifying the effects of included variables that vary little at county level over a seven-year period.

To gain insight into which of these changes is driving the differences in our results, we finally estimated two "intermediate" models between 2000 cross section and 2000-2006 fixed effects: the 2000-2006 between estimator with IV-GMM, and 2000-2006 pooled cross section estimator with IV-GMM. With results regarding heterogeneity similar between 2000 cross section and fixed effects, we focus on the "narrow>0/full/without heterogeneity" specification for this exercise. The between estimator for 2000-2006 is effectively a cross section regression conducted on the time-averaged values for all panel variables. Like 2000 cross section, the between estimator is based entirely on variation between counties and has a similar sample size, but uses some control variables at state level with clustered standard errors, and lacks state dummy variables.[25] Substantive differences in estimated results for the audit rate, gender, income, age distribution, education and income sources indicate that the move to state level controls, sampling rather than population, and lack of state dummies, is responsible for part of the differences in our results.

---

[25] Estimation would not converge with state dummies included, because of the inclusion of numerous state-level control variables.

Next, we estimate 2000-2006 pooled cross section with year dummies. This model is an unweighted average of the between estimator and the within estimator (fixed effects). While pooled cross section estimates were often intermediate to the between and within estimators as expected, results tended to be closer to those from the between estimator. This in turn indicates that unobserved heterogeneity in county characteristics is playing a large role in explaining variation in non-filing rates, and that between-estimator is likely to suffer from considerable omitted variable bias. While this may also indicate that the 2000 cross section suffers from considerable omitted variable bias, its results are still arguably of value because of its superior data and inclusion of state fixed effects.

The second contribution of the paper has been to ask whether heterogeneity has a significant effect on a compulsory rather than voluntary measure of social capital, namely tax filing compliance. We test whether heterogeneity in household income, language spoken at home, race, or religion has any effect upon non-filing rates. We find that broad and narrow non-filing rates are rising in racial fragmentation in both cross section and fixed effects analysis. Judging by our panel results, the effect of racial fragmentation may be largest on the "softest" definition of non-filing rates, which includes late-filers and non-filers owing no money. The magnitude of effect drops when late-filers are excluded and when non-filers are restricted to those estimated to have a positive balance owing. The effect attenuates to insignificance when non-filers are further restricted to those estimated to have an outstanding balance of $500 or greater. Among other dimensions of heterogeneity, non-filing rates do not appear affected by household income inequality or language fragmentation. Non-filing rates may be falling in religious membership fragmentation, but we can address the question only in year 2000 cross section estimates.

Our estimated effects of racial fragmentation on filing compliance rates are consistent with the findings of previous studies in the social capital literature that estimate the effects of racial heterogeneity on compulsory activities such as census return rates (Vigdor 2004), or on voluntary activities such as volunteering, voting, or being a member of an organization (Alesina and La Ferrara 2000, 2002; Costa and Kahn 2003a, 2003b; Putnam 2007; Clark and Kim 2012). To the extent that these findings regarding racial fragmentation persist, it seems worthwhile to explore ways to reduce racial heterogeneity's effects on both compulsory and voluntary indicators of social capital, especially given that the United States and indeed all OECD countries are growing ever more racially diverse (Putnam 2007). As discussed in Li (2010), tax agencies that emphasize common benefits from a well-functioning tax system might reduce filing non-compliance. More broadly, in increasingly diverse societies, governments that are concerned with social capital may wish to set policies that emphasize shared citizenship and identity, and values that are held in common.

References

Alesina, Alberto and Eliana La Ferrara (2000). Participation in heterogeneous communities. *The Quarterly Journal of Economics*, 115 (3), 847-904.

Alesina, Alberto and Eliana La Ferrara (2002). Who trusts others? *Journal of PublicEconomics*, 85 (2), 207-234.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg (2003). Fractionalization. *Journal of Economic Growth* 8 (2), 155-194.

Alesina, Alberto, Reza Baqir, and William Easterly (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics*, 114 (4), 1243-1284.

Allingham, Michael and Agnar Sandmo (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1 (3-4), 323-338.

Alm, James (1999). Tax compliance and administration. In W. Bartley Hildreth and James A. Richardson (Eds.), *Handbook on Taxation*. New York, NY: Marcel Dekker, Inc., 741-768.

Alm, James (2012). Measuring, explaining, and controlling tax evasion: Lessons from theory, experiments, and field studies. *International Tax and Public Finance*, 19 (1), 54-77.

Alm, James, Todd Cherry, Michael Jones, and Michael McKee, M. (2010). Taxpayer information assistance services and tax reporting behavior. *Journal of Economic Psychology*, 31 (4), 577-586.

Alm, James, Gary H. McClelland, and William D. Schulze (1999). Changing the social norm of tax compliance by voting. *Kyklos*, 52 (2), 141-172.

Andreoni, James, Brian Erard, and Jonathan Feinstein (1998). Tax compliance. *The Journal of Economic Literature*, 36 (2), 818-860.

Baum, Christopher (2008). Stata tip 63: Modeling proportions. *The Stata Journal*, 8, 2, 299-303.

Beron, Kurt, Helen Tauchen, and Ann Dryden Witte (1992). The effect of audits and socioeconomic variables on compliance. In Joel Slemrod (Ed.), *Why People Pay Taxes: Tax Compliance and Enforcement*. Ann Arbor, MI: University of Michigan Press, 67-89.

Becker, Gary S. (1968). Crime and punishment – An economic approach. *The Journal of Political Economy*, 76 (2), 169-217.

Clark, Jeremy and Bonggeun Kim (2012). The effect of neighbourhood diversity on volunteering: Evidence from New Zealand. *B.E. Journal of Economic Analysis and Policy*, 12, 1, (Topics), Article 5.

Costa, Dora and Matthew Kahn (2003a). Understanding the American decline in social capital, 1952-1998. *Kyklos*, 56 (1), 17-46.

Costa, Dora and Matthew Kahn (2003b). Civic engagement and community heterogeneity: An economist's perspective. *Perspectives on Politics*, 1 (1), 103-111.

Cowell, Frank A. (1990). *Cheating the Government: The Economics of Evasion*. Cambridge, MA: The MIT Press.

Cowell, Frank A. and James P. F. Gordon (1988). Unwillingness to pay. *Journal of Public Economics*, 36 (3), 305-321.

Dubin, Jeffrey A., Michael Graetz, and Louis L. Wilde (1990). The effect of audit rates on the federal individual income tax, 1977-1986. *National Tax Journal*, 43 (4), 395-409.

Elster, Jon (1989). *The Cement of Society – A Study of Social Order*. Cambridge, UK: Cambridge University Press.

Fortin, Bernard, Guy Lacroix, and Marie-Claire Villeval (2007). Tax evasion and social interactions. *Journal of Public Economics*, 91 (8), 2089-2112.

Gordon, James P. F. (1989). Individual morality and reputation costs as deterrents to tax evasion. *European Economic Review*, 33 (4), 797-805.

Gneezy, Uri and Also Rustichini (2000). A fine is a price. *Journal of Legal Studies*, 29 (1, Part 1), 1-18.

Erard, Brian and Chih-Chin Ho (2001). Searching for ghosts: who are the non-filers and how much tax do they owe? *Journal of Public Economics*, 81 (1), 25-50.

Gentry, William and Matthew Kahn (2009). Understanding spatial variation in tax sheltering: The role of demographics, ideology, and taxes. *International Regional Science Review*, 32 (3), 400-423.

Gustavsson, Magnus and Henrik Jordahl (2008). Inequality and trust in Sweden: Some inequalities are more harmful than others. *Journal of Public Economics*, 92, 348-365.

Hisnanick, John and Annette Rogers (undated). Household income inequality measures based on the ACS Data: 2000-2005. Housing and Household Economic Statistics Division, U.S. Census Bureau. Washington, D.C.

Internal Revenue Service (2012) http://www.irs.gov/pub/newsroom/overview_tax_gap_2006.pdf

Kim, Youngse (2003). Income distribution and equilibrium multiplicity in a stigma-based model of tax evasion. *Journal of Public Economics*, 87 (9), 1591-1616.

Kleven, Henrik J. Martin B. Knudsen, Claus T. Kreiner, Soren Pedersen, and Emmanuel Saez (2011). Unwilling or unable to cheat? Evidence from a randomized tax audit experiment in Denmark. *Econometrica*, 79 (3), 651-692.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny (1999). The quality of government. *The Journal of Law, Economics and Organization*, 15, 222-279.

Leigh, Andrew (2006). Trust, inequality and ethnic heterogeneity. *The Economic Record*, 82, 268-280.

Li, Sherry Xin (2010). Social identities, ethnic diversity, and tax morale. *Public Finance Review*, 38 (2), 146-177.

Luttmer, Erzo (2001). Group loyalty and the taste for redistribution. *The Journal of Political Economy*, 109 (3), 500-528.

Miguel, Edward and Mary Kay Gugerty (2005). Ethnic diversity, social sanctions, and public goods in Kenya. *Journal of Public Economics*, 89 (11-12), 2325-2368.

Myles, Gareth D. and Robin A. Naylor (1996). A model of tax evasion with group conformity and social customs. *European Journal of Political Economy*, 12 (1), 49-66.

Plumley (1996) ?

Poterba, James M. (1997). Demographic structure and the political economy of public education. *Journal of Policy Analysis and Management*, 16 (1), 48-66.

Putnam, Robert (2007). E pluribus unnum: Diversity and community in the twenty-first century – The 2006 Johan Skytte prize lecture. *Scandinavian Political Studies*, 30, 137-174.

Sandmo, Agnar (2005). The theory of tax evasion: A retrospective view. *National Tax Journal*, 58 (4), 643-663.

Slemrod, Joel and Shlomo Yitzhaki (2002). Tax avoidance, evasion, and administration. In Alan J. Auerbach and Martin Feldstein (Eds.), *Handbook of Public Economics*. Amsterdam, London, and New York: Elsevier, 1423-1470.

Traxler, Christian (2010). Social corms and conditional cooperative taxpayers. *European Journal of Political Economy*, 26 (1), 89-103.

Vigdor, Jacob (2004). Community composition and collective action: Analyzing initial mail responses to the 2000 census. *The Review of Economics and Statistics*, 86 (1), 303-312.

Figure 1: County Non-filing Rates by Lagged Enforcement Variables, Tax Year 2000 (N = 3027)
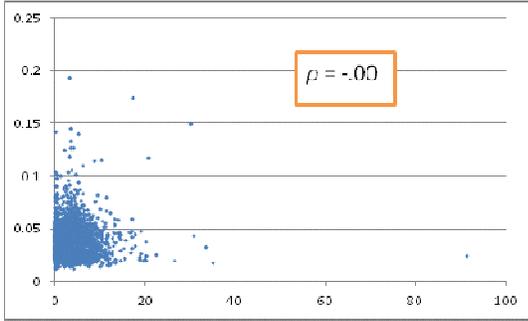
A. Broad Non-filing Rate as a Function of
    Average Penalty Rate for Non-Filers
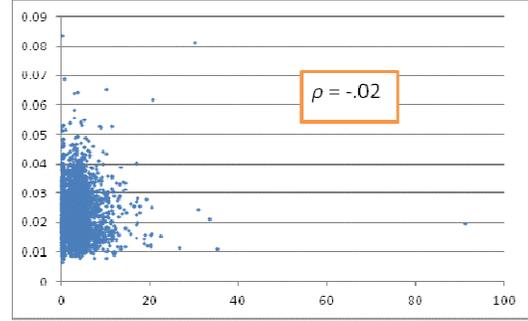
B. Narrow Non-filing Rate as a Function of
    Average Penalty Rate for Non-Filers
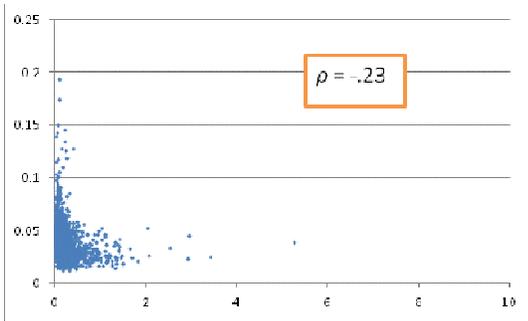
C. Broad Non-filing Rate as a Function of
    Enforcement Index Against Identified Non-Filers
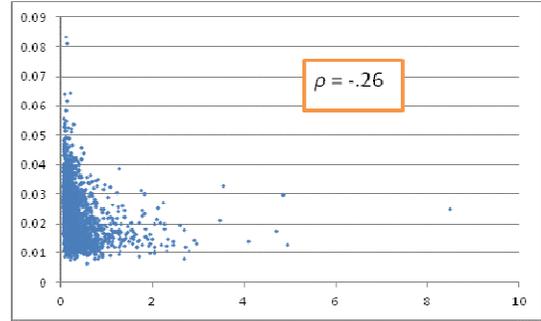
D. Narrow Non-filing Rate as a Function of
    Enforcement Index Against Identified Non-Filers
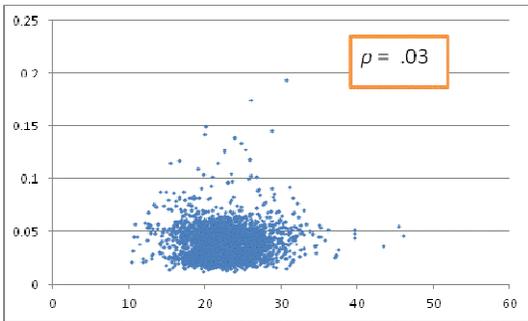
E. Broad Non-filing Rate as a Function of
    Average Penalty Rate for Tax Filers
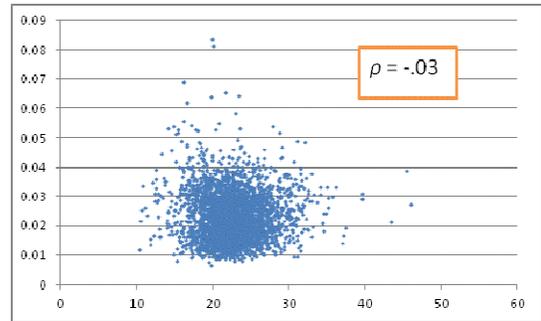
F. Narrow Non-filing Rate as a Function of
    Average Penalty Rate for Tax Filers
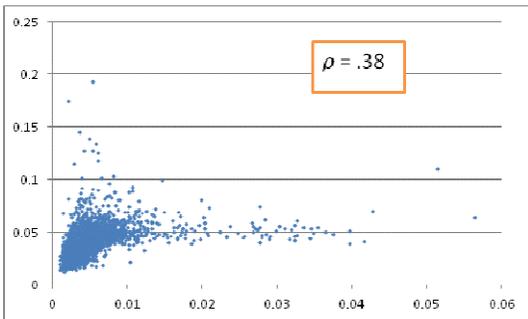
G. Broad Non-filing Rate as a Function of
    Audit Rate of Filers
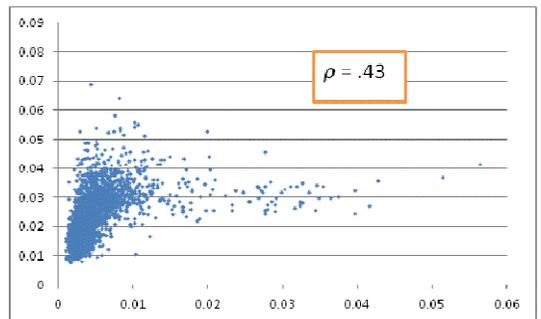
H. Narrow Non-filing Rate as a Function of
    Audit Rate of Filers

Figure 2: County Non-filing Rates by Fragmentation for Tax Year 2000 (N = 3027)
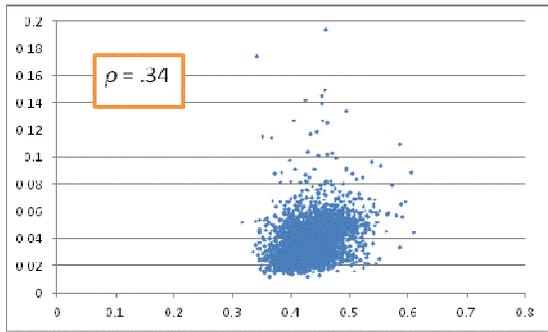
A. Broad Non-filing Rate as a Function of Household Income Inequality (Gini)
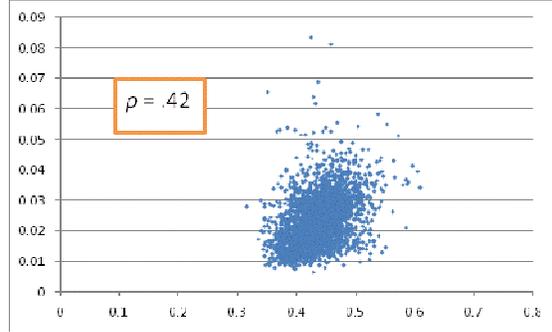
B. Narrow Non-filing Rate as a Function of Household Income Inequality (Gini)
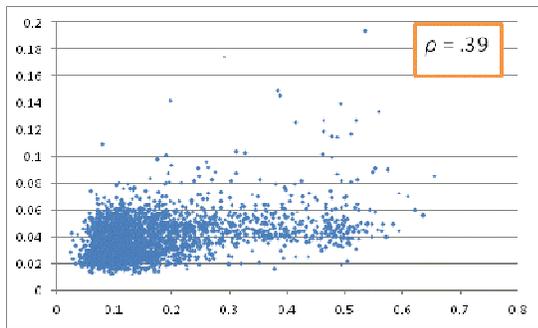
C. Broad Non-filing Rate as a Function of Fragmentation by Home Language
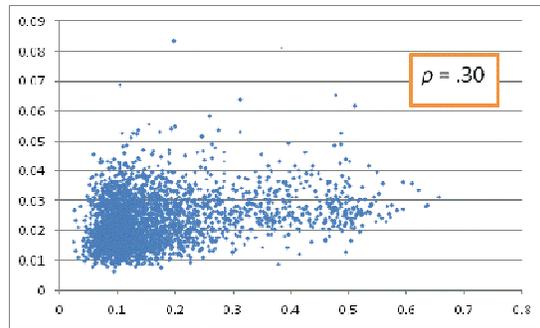
D. Narrow Non-filing Rate as a Function of Fragmentation by Home Language
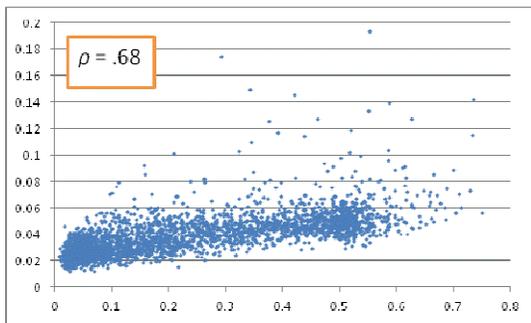
E. Broad Non-filing Rate as a Function of Fragmentation by Race
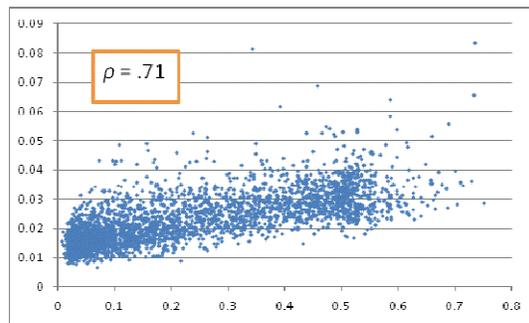
F. Narrow Non-filing Rate as a Function of Fragmentation by Race
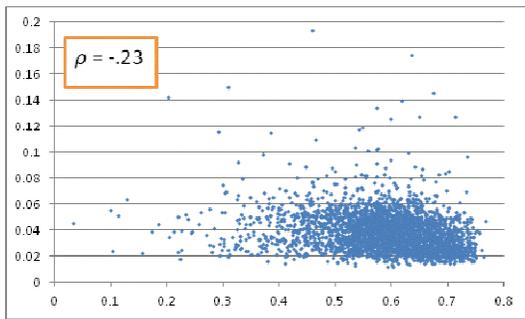
G. Broad Non-filing Rate as a Function of Fragmentation by Religion
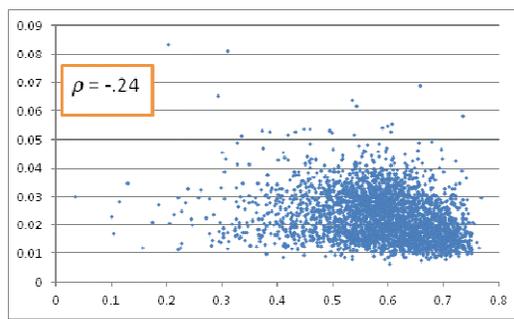
H. Narrow Non-filing Rate as a Function of Fragmentation by Religion

Table 1:
Descriptive Statistics for Variables in Tax Year 2000 Cross Section Regressions (All at County Level)

Variable	N	Mean	Odds	St.Dev.	Min	Max
Broad Non-filing Rate	3027	.0392	.0408	.0154	.0122	.1938
Narrow Non-filing Rate >$0	3027	.0232	.0238	.0085	.0068	.0834
Narrow Non-filing Rate >$500	3022	.0129	.0131	.0055	.0033	.0572
Enforcement Index Broad	**	**		**	**	**
Enforcement Index Narrow >$0	**	**		**	**	**
Enforcement Index Narrow >$500	**	**		**	**	**
Audit Rate Filers	2638	.0050		.0046	.0009	.0564
Penalty Rate Non-Filers	3133	3.491		3.812	0	91.28
Penalty Rate Filers	3141	22.67		3.989	10.29	55.38
NN Enforcement Index Broad	**	**		**	**	**
NN Enforcement Index Narrow >$0	**	**		**	**	**
NN Enforcement Index Narrow >$500	**	**		**	**	**
NN Audit Rate Filers	2639	.0049		.0042	.0009	.0564
Information Docs. Per Filed Return	3141	8.660		7.125	2.636	170.6
Information Docs. Types /23	3141	.7561		.1963	0	1
Percent Returns Seeking EITC	3138	.1826		.0830	.028	.584
Share Male	3140	.4962		.0222	0	.6758
Real Median HH Income/10,000	3138	3.637		.9006	1.523	9.121
Education < High School	3140	.2262		.0876	.0304	.6530
Education High School Diploma	3140	.3469		.0655	.1093	.5325
Education Some College/Assoc Degree	3140	.2617		.0567	.0952	.4489
Education Bachelor's Degree	3140	.1097		.0492	0	.4002
Education Masters/PhD/Prof Degree	3140	.0555		.0329	0	.3603
Share Age < 15	3139	.2063		.0283	.0205	.3809
Share Age 15-19	3139	.0754		.0128	0	.2438
Share Age 20-24	3139	.0602		.0253	.0137	.2918
Share Age 25-44	3139	.2749		.0316	.1452	.4814
Share Age 45-64	3139	.2359		.0277	.0537	.4589
Share Age 65 Plus	3139	.1473		.0419	.0179	.3466
Share of Household Income from:						
Wages and Salaries	3140	.6859		.0736	.3844	.8806
Self-employment	3140	.0762		.0368	0	.3896
Interest/Dividends/Net Rental	3140	.0669		.0267	.0051	.2411
Social Security	3140	.0767		.0247	.0052	.1882
Supplemental Social Security	3140	.0074		.0053	0	.0603
Public Assistance	3140	.0020		.0024	0	.0639
Retirement Funds	3140	.0605		.0238	.0030	.2524
Other Income	3140	.0244		.0072	0	.0784

Table 1: (Cont'd)
Descriptive Statistics for County Level Variables in 2000 Cross Section Regression

Variable	N	Mean	St.Dev.	Min	Max
Never Married	3140	.2248	.0562	.0902	.5613
Currently Married (not Separated)	3140	.5854	.0592	.1973	.8759
Widowed/Divorced/Separated	3140	.1898	.0302	.0338	.6054
Share English at Home	3140	.8940	.1154	.0444	.9881
Share Spanish at Home	3140	.0624	.1037	0	.9497
Share Other Language at Home	3140	.0436	.0547	0	.7193
Median Years at Residence	3140	7.249	2.195	2	18
Share Foreign Born	3140	.0347	.0486	0	.5094
Median Home Value/10,000	3140	8.458	4.773	0	100.0
Share White Non-Hispanic	3139	.8221	.1911	.0210	.9974
Share White Hispanic	3139	.0589	.1181	0	.9693
Share Black Non-Hispanic	3139	.0878	.1452	0	.8601
Share Black Hispanic	3139	.0015	.0033	0	.1104
Share American Indian/Alaskan Native	3139	.0198	.0778	0	.9503
Share Asian	3139	.0091	.0247	0	.5860
Share Hawaiian/Pacific Islander	3139	.0009	.0090	0	.4345
Share Mainline Protestant	3139	.1413	.1125	0	.8188
Share Evangelical Protestant	3139	.2266	.1675	0	.9833
Share Catholic	3139	.1363	.1471	0	.8999
Share Orthodox	3139	.0009	.0060	0	.1914
Share Other Religions	3139	.0234	.0846	0	.9157
Share Unclaimed Adherents	3139	.4714	.1811	0	.9818
Home Ownership Rate	3140	.7394	.0777	0	.8954
Average Household Size	3140	2.538	.2040	1.28	4.38
Unemployment Rate	3140	.0476	.0262	0	.2760
Share Civilian Pop in Labor Force	3139	.7359	.1123	0	1.563
Share Poverty (x 100)	3139	13.29	5.595	0	42.20
Fragmentation by:					
Household Income (Gini)	3140	.4342	.0388	.3152	.6085
Language Spoken at Home	3140	.1679	.1138	.0235	.6554
Race	3139	.2342	.1856	.0053	.7493
Religion	3139	.5850	.0985	.0328	.7681

Table 2:
Descriptive Statistics for County and State Level Variables in 2000-2006 Panel Regression

Variable	N	Mean	Odds	St.Dev.	Min	Max	State/County
Broad Non-filing Rate	21132	.0354	.0367	.0143	.0067	.2051	County
Narrow Non-filing Rate >$0	21127	.0204	.0208	.0084	.0041	.0834	County
Narrow Non-filing Rate >$500	21111	.0143	.0145	.0063	.0028	.0710	County
Enforcement Index Broad	20540	**		**	**	**	County
Enforcement Index Narrow >$0	20506	**		**	**	**	County
Enforcement Index Narrow >$500	20435	**		**	**	**	County
Audit Rate Filers	19012	.0058		.0047	.0006	.0869	County
Penalty Rate Non-Filers	21911	3.654		3.800	0	100.0	County
Penalty Rate Filers	21992	20.42		4.237	.0121	66.72	County
NN Enforcement Index Broad	20395	**		**	**	**	County
NN Enforcement Index Narrow >$0	20361	**		**	**	**	County
NN Enforcement Index Narrow >$500	20286	**		**	**	**	County
NN Audit Rate Filers	19015	.0059		.0047	.0009	.0869	County
Percent Returns Seeking EITC	21973	.1907		.0855	.0270	.6060	County
Share Male	21982	.4966		.0208	0	.7010	County
Real Median HH Income/10,000	21987	3.370		.8658	0	8.824	County
Education < High School	21987	.1650		.0431	.0858	.2590	State
Education High School Diploma	21987	.3109		.0399	.1932	.4269	State
Education Some College/Assoc Degree	21987	.2771		.0321	.1685	.3839	State
Education Bachelor's Degree	21987	.1612		.0261	.0891	.2312	State
Education Masters/PhD/Prof Degree	21987	.0857		.0191	.0543	.2599	State
Share Age < 15	21980	.1974		.0294	0	.3834	County
Share Age 15-19	21980	.0741		.0120	0	.2438	County
Share Age 20-24	21980	.0680		.0225	0	.3339	County
Share Age 25-44	21980	.2631		.0351	.1271	.4814	County
Share Age 45-64	21980	.2491		.0303	.0492	.6396	County
Share Age 65 Plus	21980	.1483		.0414	.0149	.3522	County
Share of Household Income from:							
Wages and Salaries	21987	.7489		.0287	.6442	.8006	State
Self-employment	21987	.0656		.0155	.0379	.1311	State
Interest/Dividends/Net Rental	21987	.0486		.0103	.0286	.0969	State
Social Security	21987	.0619		.0131	.0230	.1085	State
Supplemental Social Security	21987	.0046		.0018	.0017	.0119	State
Public Assistance	21987	.0011		.0005	.0003	.0052	State
Retirement Funds	21987	.0496		.0089	.0308	.0770	State
Other Income	21987	.0198		.0030	.0116	.0300	State

Table 2 (Cont'd):
Descriptive Statistics for County and State Level Variables in 2000-2006 Panel Regression

Variable	N	Mean	St.Dev.	Min	Max	State or County Level?
Never Married	21987	.2648	.0254	.2136	.5402	State
Currently Married (not Separated)	21987	.5478	.0262	.2695	.6086	State
Widowed/Divorced/Separated	21987	.1874	.0164	.1310	.2327	State
Speak English at Home	21987	.8787	.0963	.5747	.9791	State
Speak Spanish at Home	21987	.0739	.0812	.0060	.2923	State
Speak Other Language at Home	21987	.0474	.0293	.0087	.2486	State
Same Residence Year Ago	21987	.8372	.0222	.7617	.8949	State
Share Foreign Born	21987	.0720	.0562	.0083	.2724	State
Real Median Home Value/10000	21987	11.647	4.771	6.968	4.427	State
Share White Non-Hispanic	21980	.8153	.1932	.0202	.9986	County
Share White Hispanic	21980	.0635	.1208	0	.9693	County
Share Black Non-Hispanic	21980	.0885	.1454	0	.8674	County
Share Black Hispanic	21980	.0017	.0034	0	.1180	County
Share American Indian/Alaskan Native	21980	.0201	.0780	0	.9507	County
Share Asian	21980	.0101	.0255	0	.5860	County
Share Hawaiian/Pacific Islander	21980	.0009	.0085	0	.4470	County
Home Ownership Rate	21987	.6857	.0395	.3863	.7665	State
Average Household Size	21980	2.135	.3423	.5015	5.005	County
Share Poverty (x100)	21977	14.09	5.763	0	51.0	State
Unemployment Rate	21971	.0549	.0251	0	.276	County
Share Civilian Pop in Labor Force	21980	.7386	.1117	0	1.636	County
Average State Income Tax Rate	21987	.0195	.0107	0	.0432	State
Fragmentation by:						
Household Income (Gini)	21987	.4443	.0191	.3841	.5448	State
Language Spoken at Home	21987	.2034	.1324	.0412	.5692	State
Race	21980	.2431	.1863	.0028	.7550	County

Table 3: Determinants of non-filing rates: IV-GMM cross section regression for Tax Year 2000[1]

Variable	(1) Broad NF Sparse	(2) Broad NF Full	(3) Broad NF+Het Sparse	(4) Broad NF+Het Full	(5) Narrow NF Sparse	(6) Narrow NF Full	(7) Narrow NF+Het Sparse	(8) Narrow NF+Het Full
Intercept	-1.28* (.713)	-.624 (.894)	-2.28*** (.716)	-1.27 (.866)	-1.39** (.676)	-.909 (.787)	-2.58*** (.663)	-1.85** (.780)
Penalty Rate Non-filers	.002 (.002)	.001 (.002)	.003 (.002)	.001 (.001)	.002 (.001)	.000 (.001)	.002 (.001)	.001 (.001)
Enforcement Index Non-filers	-.753*** (.170)	-.541*** (.186)	-.722*** (.168)	-.535*** (.183)	-.411*** (.094)	-.272*** (.102)	-.389*** (.092)	-.267*** (.101)
Penalty Rate Filers	.001 (.002)	-.002 (.002)	.001 (.002)	-.002 (.002)	.000 (.002)	-.002 (.002)	.000 (.002)	-.002 (.002)
Audit Rate Filers	-.781 (1.68)	.316 (1.72)	-.323 (1.66)	.330 (1.69)	.507 (1.72)	1.74 (1.77)	.986 (1.70)	1.68 (1.73)
Share Male	.394 (.410)	.943** (.399)	-.031 (.417)	.647 (.396)	.104 (.399)	.572 (.391)	-.413 (.406)	.312 (.392)
Median HH Income	-.286** (.114)	-.461*** (.007)	-.139 (.117)	-.332*** (.123)	-.441*** (.090)	-.538*** (.091)	-.267*** (.090)	-.383*** (.087)
Age Shares 15-19	-4.14** (1.18)	-3.71*** (1.08)	-3.85*** (1.10)	-3.65*** (1.05)	-3.61*** (1.20)	-3.48*** (1.09)	-3.26*** (1.11)	-3.35*** (1.07)
20-24	.091 (.508)	-1.25* (.658)	.675 (.523)	-1.03 (.668)	-.132 (.498)	-1.32** (.640)	-.568 (.503)	-1.06 (.657)
25-44	-1.53** (.758)	-1.82*** (.654)	-.946 (.734)	-1.23** (.621)	-1.09 (.710)	-1.57** (.624)	-.380 (.685)	-.907 (.594)
45-64	1.39** (.559)	.960* (.532)	2.02*** (.551)	1.46*** (.503)	1.22** (.537)	.865* (.505)	1.98*** (.518)	1.43*** (.473)
65 Plus	-2.09*** (.575)	-2.53*** (.590)	-1.60*** (.555)	-1.63*** (.556)	-1.41*** (.551)	-2.26*** (.568)	-.821 (.528)	-1.33** (.535)
Share White H	.564*** (.118)	.603 (.405)	.335*** (.128)	-.362 (.481)	.544*** (.101)	.913*** (.345)	.272** (.110)	-.238 (.395)
Black NH	.924*** (.057)	.717*** (.077)	.631*** (.082)	.321*** (.105)	.820*** (.054)	.599*** (.074)	.468*** (.079)	.182* (.106)
Black H	5.90*** (2.11)	.770 (1.40)	4.38*** (1.65)	-1.59 (1.47)	3.03** (1.51)	-.435 (1.23)	1.22 (1.04)	-3.21** (1.41)
AmInd/AlaskNat	1.20*** (.136)	.946*** (.131)	.923*** (.150)	.592*** (.156)	1.17*** (.113)	.881*** (.123)	.838*** (.122)	.512*** (.151)
Asian	.806*** (.282)	-.187 (.358)	.139 (.294)	-.780** (.362)	.845*** (.261)	.120 (.331)	.044 (.266)	-.503 (.323)
Hawaiian/PI	8.95*** (3.32)	7.50** (3.72)	7.57** (3.35)	5.80 (3.62)	8.11*** (2.86)	6.72** (3.06)	6.46** (2.87)	5.31* (3.04)
Share in Civ. Labor Force	-.152** (.067)	-.194*** (.065)	-.153** (.066)	-.184*** (.065)	-.183*** (.063)	-.240*** (.061)	-.185*** (.062)	-.223*** (.060)
Average HH Size	-1.07*** (.213)	-1.11*** (.274)	-.729*** (.222)	-.756*** (.255)	-1.26*** (.190)	-1.21*** (.214)	-.854*** (.192)	-.798*** (.200)
Unemployment Rate	-.584* (.305)	-.260 (.325)	-.503* (.300)	-.156 (.315)	-.657*** (.244)	-.343 (.256)	-.549** (.233)	-.216 (.245)
HH Inc X Average HH Size	.266*** (.061)	.314*** (.067)	.186*** (.062)	.235*** (.063)	.335*** (.051)	.358*** (.050)	.240*** (.051)	.267*** (.047)
(HH Inc X Average HH Size)^2	-.004*** (.001)	-.004*** (.001)	-.003*** (.001)	-.003*** (.001)	-.005*** (.001)	-.005*** (.001)	-.004*** (.001)	-.004*** (.001)
Share Poverty	.026*** (.004)	.014*** (.004)	.025*** (.004)	.014*** (.004)	.031*** (.004)	.020*** (.004)	.030*** (.003)	.018*** (.004)

Table 3 Cont'd: Determinants of non-filing rates: IV-GMM cross section regression for Tax Year 2000[1]

Variable	(1) Broad NF Sparse	(2) Broad NF Full	(3) Broad NF+Het Sparse	(4) Broad NF+Het Full	(5) Narrow NF Sparse	(6) Narrow NF Full	(7) Narrow NF+Het Sparse	(8) Narrow NF+Het Full
Education < High School		.080 (.181)		.013 (.184)		.215 (.175)		.099 (.171)
Some College		.640*** (.182)		.592*** (.182)		.591*** (.182)		.524*** (.182)
Bachelors		-.385 (.342)		-.351 (.320)		.084 (.264)		.033 (.235)
Masters/PhD		.900* (.480)		.846* (.483)		.491* (.261)		.402 (.253)
Share Inc. Self-employment		1.82*** (.299)		1.62*** (.287)		1.68*** (.272)		1.46*** (.267)
Interest/Div/Rent		.033 (.286)		.058 (.282)		-.032 (.268)		-.089 (.274)
Social Security		.891 (.735)		.360 (.755)		1.82*** (.664)		1.47** (.690)
Supplemental SS		-.690 (2.75)		.079 (2.66)		-4.07* (2.43)		-3.08 (2.36)
Public Assistance		15.4*** (5.72)		14.5*** (5.46)		16.0*** (4.82)		15.5*** (.190)
Retirement Funds		.010 (.291)		.035 (.294)		-.002 (.270)		.021 (.272)
Other Sources		1.29 (1.17)		1.43 (1.16)		1.17 (1.01)		1.43 (.978)
Share Foreign Born		1.11*** (.244)		.945*** (.238)		.921*** (.244)		.723*** (.237)
Speak Spanish at Home		-.332 (.490)		.425 (.532)		-.738* (.410)		.199 (.428)
Speak Other Lang at Home		-.122 (.117)		-.250 (.208)		-.186* (.109)		-.363* (.186)
Share Married		-1.32*** (.294)		-1.53*** (.323)		-1.42*** (.293)		-1.63*** (.328)
Share Wid/Div/Separated		1.01* (.589)		.685 (.615)		.979* (.561)		.613 (.590)
Mainline Protestant		-.070 (.092)		-.002 (.102)		-.076 (.082)		-.017 (.092)
Evangelical Protestant		-.095** (.038)		-.102*** (.039)		-.071** (.035)		-.090** (.023)
Catholic		-.106* (.063)		-.091 (.071)		-.089* (.053)		-.089 (.059)
Orthodox		.371 (1.12)		.934 (1.13)		.851 (1.10)		1.24 (1.15)
Other Religion		-.035 (.097)		-.101 (.093)		.007 (.095)		-.076 (.091)
Median Years Residence		-.027*** (.004)		-.027*** (.004)		-.023*** (.004)		-.023*** (.004)
Median Home Value		.007*** (.002)		.007*** (.002)		.006*** (.002)		.006*** (.002)
Share HH Owner Occup.		.355** (.146)		.325** (.146)		.343*** (.129)		.314** (.128)
Heterogeneity by: HH Inc.				-.093 (.265)				.163 (.231)
Language				.103 (.149)				.165 (.134)
Race			.355*** (.070)	.400*** (.078)			.426*** (.065)	.423*** (.072)
Religion				-.170** (.072)				-.123* (.065)
Centered R²	.747	.815	.756	.821	.756	.821	.767	.827
N	2276	2275	2276	2275	2276	2275	2276	2275

[1] All estimations are run on Stata 12.0 using ivreg2 and two step GMM. Numbers in parentheses are robust standard errors. *, **, and *** refer to two tailed significance at the 10, 5, and 1 percent levels, respectively. State dummies are included, but not reported. Enforcement Index Non-filers and Audit Rate Filers are instrumented.

Table 4: Determinants of non-filing rates: IV-GMM fixed effects regression for Tax Years 2000-2006[1]

Variable	(1) Broad NF Sparse	(2) Broad NF Full	(3) Broad NF+Het Sparse	(4) Broad NF+Het Full	(5) Narrow NF Sparse	(6) Narrow NF Full	(7) Narrow NF+Het Sparse	(8) Narrow NF+Het Full
Penalty Rate Non-filers	.000 (.000)	.000 (.000)	.000 (.000)	-.000 (.000)	-.000 (.000)	-.000 (.000)	-.000 (.000)	-.000* (.000)
Enforcement Index Non-filers	-.037*** (.011)	-.028* (.015)	-.038*** (.011)	-.028* (.015)	-.009 (.007)	.000 (.012)	-.010 (.007)	-.000 (.012)
Penalty Rate Filers	.000 (.000)	.000 (.001)	.000 (.000)	.000 (.001)	.000 (.000)	.000 (.001)	.001 (.000)	.000 (.001)
Audit Rate Filers	2.15*** (.391)	2.96*** (.836)	2.19*** (.394)	2.96*** (.846)	3.35*** (.397)	2.57*** (.707)	3.37*** (.399)	2.58*** (.723)
Share Male	-.871*** (.258)	-.804** (.358)	-.869*** (.264)	-.810** (.384)	-.644** (.399)	-.742** (.343)	-.643** (.291)	-.747** (.354)
Real Median HH Income	-.008 (.098)	.054 (.101)	-.010 (.098)	.049 (.101)	-.047 (.045)	.038 (.091)	-.048 (.045)	.034 (.065)
Age Shares 15-19	-.014 (.409)	-.006 (.802)	-.171 (.411)	-.227 (.834)	.197 (.364)	.313 (.694)	.094 (.368)	.166 (.695)
20-24	.970*** (.192)	.860* (.461)	.854*** (.192)	.739 (.459)	.562*** (.175)	.758* (.412)	.486*** (.175)	.671* (.404)
25-44	1.01*** (.203)	.940* (.485)	.873*** (.203)	.786 (.490)	.533*** (.191)	.633 (.391)	.444** (.192)	.529 (.386)
45-64	.865*** (.235)	.848 (.576)	.786*** (.234)	.767 (.569)	.116 (.206)	.409 (.423)	.065 (.206)	.358 (.417)
65 Plus	.884*** (.308)	1.27* (.727)	.840*** (.308)	1.21* (.728)	1.29*** (.273)	1.44** (.683)	1.26*** (.273)	1.39** (.686)
Share White H	1.38*** (.171)	1.05*** (.318)	.767*** (.218)	.420 (.324)	1.76*** (.177)	1.70*** (.416)	1.37*** (.214)	1.24*** (.395)
Black NH	.964*** (.148)	.872*** (.285)	.437** (.174)	.289 (.288)	.675*** (.139)	.568** (.237)	.331** (.162)	.164 (.291)
Black H	-10.1*** (3.17)	-9.06* (5.39)	-10.7*** (3.17)	-9.12* (5.19)	-3.06 (2.22)	-2.41 (2.97)	-3.42 (2.22)	-2.63 (3.01)
AmInd/AlaskNat	-.329 (.885)	-.545 (1.10)	-.789 (.882)	-1.02 (1.06)	.517 (.643)	.406 (.589)	.212 (.644)	.059 (.577)
Asian	.472 (.391)	-.239 (.847)	-.396 (.396)	-1.21 (.849)	-.448 (.324)	.070 (.642)	-1.02*** (.357)	-.622 (.657)
Hawaiian/PI	-1.60 (1.70)	-.908 (1.64)	-3.21* (1.74)	-2.81* (1.60)	-1.05 (1.77)	-.270 (2.00)	-2.12 (1.80)	-1.60 (1.72)
Share Civilian Labor Force	.025 (.019)	.031 (.028)	.024 (.019)	.031 (.028)	-.001 (.019)	-.013 (.025)	-.002 (.019)	-.013 (.026)
Average HH Size	-.027 (.146)	.107 (.159)	-.034 (.146)	.097 (.158)	-.038 (.080)	.081 (.131)	-.043 (.080)	.074 (.130)
Unemployment Rate	.416*** (.090)	.574*** (.161)	.385*** (.090)	.549*** (.163)	.342*** (.078)	.360** (.166)	.322*** (.079)	.343** (.170)
HH Inc X Average HH Size	.025 (.055)	-.010 (.059)	.026 (.055)	-.007 (.059)	.053** (.027)	.013 (.045)	.053** (.027)	.015 (.044)
(HH Inc X Average HH Size)^2	-.001* (.001)	-.001 (.001)	-.001* (.001)	-.001 (.001)	-.002*** (.000)	-.002** (.001)	-.002*** (.000)	-.002** (.001)
Share Poverty	-.001 (.001)	-.000 (.001)	-.001 (.001)	-.001 (.001)	.001 (.001)	.001 (.001)	.001 (.001)	.001 (.001)

Table 4 Cont'd: Determinants of non-filing rates: IV-GMM fixed effects regression for Tax Years 2000-2006[1]

Variable	(1) Broad NF Sparse	(2) Broad NF Full	(3) Broad NF+Het Sparse	(4) Broad NF+Het Full	(5) Narrow NF Sparse	(6) Narrow NF Full	(7) Narrow NF+Het Sparse	(8) Narrow NF+Het Full
Education < High School[2]		-.727 (.546)		-.784 (.556)		-1.18*** (.428)		-1.22*** (.437)
Some College[2]		-.543 (.479)		-.666 (.498)		-.363 (.487)		-.433 (.499)
Bachelors[2]		-.473 (.933)		-.594 (.927)		-.785 (.570)		-.839 (.596)
Masters/PhD[2]		-.430 (.480)		-.782 (.857)		-.611 (.866)		-.799 (.894)
Share Inc.Self-employment[2]		.682 (.486)		.701 (.490)		.817** (.350)		.825** (.362)
Interest/Div/Rent[2]		-.815 (.550)		-.812 (.526)		-.459 (.630)		-.466 (.597)
Social Security[2]		-4.09** (1.72)		-4.16** (1.70)		-1.68 (1.79)		-1.71 (1.78)
Supplemental SS[2]		-6.31 (9.09)		-6.33 (9.10)		2.01 (6.99)		1.83 (7.07)
Public Assistance[2]		64.1*** (24.0)		61.9*** (23.3)		31.9*** (12.2)		30.9*** (11.5)
Retirement Funds[2]		-.751 (1.37)		-.611 (1.34)		-1.32 (1.35)		-1.24 (1.34)
Other Sources[2]		3.38 (2.11)		3.39 (2.08)		4.81** (1.88)		4.80*** (1.87)
Share Foreign Born[2]		-3.57** (1.64)		-3.50** (1.59)		1.05 (.896)		1.08 (.873)
Speak Spanish at Home[2]		3.31*** (.490)		1.88 (1.43)		.860 (.757)		.226 (1.19)
Speak Other Lan at Home[2]		.088 (1.41)		-1.60 (2.17)		-1.22 (.899)		-1.96 (1.70)
Share Married[2]		-1.22* (.626)		-1.29** (.643)		-1.01* (.574)		-1.05* (.584)
Share Wid/Div/Separated[2]		-1.34 (.589)		-1.36 (.909)		-.524 (.765)		-.548 (.782)
Average State Income Tax Rate[2]		5.59** (2.20)		5.55** (2.22)		2.73 (2.14)		2.73 (2.12)
Share Same House Yr Ago[2]		-1.44*** (.445)		-1.48*** (.445)		-1.26*** (.481)		-1.28*** (.489)
Real Median Home Value[2]		.002 (.001)		.002 (.002)		-.002 (.001)		-.001 (.001)
Share HH Owner Occup.[2]		1.20** (.527)		1.30** (.523)		.809 (.505)		.864* (.503)
Heterogeneity by: HH Inc.[2]				.125 (.455)				.087 (.393)
Language[2]				1.04 (.988)				.460 (.806)
Race			.670*** (.148)	.740*** (.230)			.437*** (.136)	.512** (.251)
Centered R[2]	.025	.090	.027	.094	.029	.064	.030	.066
N	16470	16470	16470	16470	16470	16470	16470	16470

[1] All estimations are run on Stata 12.0. Numbers in parentheses are robust standard errors. Standard errors are clustered to the state level in models (2), (4), (6) and (8). *, **, and *** refer to two tailed significance at the 10, 5 and 1 percent levels, respectively. Year dummies are included but not reported. Enforcement Index Non-filers and Audit Rate Filers are instrumented.

[2] These variables are available only at the state, rather than the county, level.

Table 5: Determinants of non-filing rates of those owing at least $500: IV-GMM regressions[1]

Variable	(1) Cross Section Sparse	(2) Cross Section Full	(3) Cross Section+Het Sparse	(4) Cross Section+Het Full	(5) Panel Sparse	(6) Panel Full	(7) Panel+Het Sparse	(8) Panel+Het Full
Intercept	-2.46***(.801)	-1.33 (1.00)	-3.56***(.802)	-.569 (.823)				
Penalty Rate Non-filers	.003*(.002)	.002 (.002)	.003*(.002)	.002 (.002)	-.000 (.000)	-.000 (.000)	-.000 (.000)	-.000 (.000)
Enforcement Index Non-filers	-.297***(.059)	-.219***(.066)	-.285***(.059)	-.217***(.065)	.008 (.005)	.017*(.010)	.008 (.005)	.017*(.010)
Penalty Rate Filers	-.000 (.002)	-.003 (.002)	-.000 (.002)	-.004*(.002)	.000 (.000)	.000 (.001)	.001 (.000)	.000 (.001)
Audit Rate Filers	.118 (1.75)	1.69 (1.87)	.634 (1.75)	1.64 (1.83)	4.47***(.441)	3.54***(.703)	4.48***(.442)	3.52***(.718)
Share Male	.231 (.478)	.655 (.470)	-.250 (.489)	.390 (.470)	.061 (.281)	-.063 (.351)	.063 (.282)	-.061 (.356)
Median HH Income	-.352***(.109)	-.475***(.116)	-.190*(.113)	-.320***(.111)	-.117**(.048)	-.013 (.063)	-.118**(.048)	-.014 (.063)
Age Shares 15-19	-4.41***(1.46)	-4.23***(1.39)	-4.11***(1.38)	-4.11***(1.36)	-.622 (.447)	-.137 (.558)	-.682 (.452)	-.208 (.565)
20-24	.219 (.597)	-1.25 (.796)	.874 (.610)	-1.00 (.821)	-.140 (.176)	.311 (.342)	-.189 (.177)	.259 (.348)
25-44	-1.59*(.873)	-2.13***(.778)	-.937 (.852)	-1.48**(.747)	-.220 (.190)	.098 (.258)	-.276 (.192)	.037 (.265)
45-64	1.52**(.641)	1.14*(.626)	2.22***(.628)	1.67***(.595)	-1.08***(.210)	-.552*(.323)	-1.11***(.210)	-.568*(.323)
65 Plus	-1.58**(.667)	-2.02***(.716)	-1.03 (.648)	-1.08 (.689)	.512*(.280)	.741 (.574)	.493*(.280)	.715 (.578)
Share White H	.599***(.118)	1.08**(.429)	.346***(.130)	-.040 (.493)	.938***(.183)	1.03***(.321)	.688***(.221)	.724*(.371)
Black NH	.900***(.067)	.665***(.092)	.574***(.094)	.261**(.128)	.627***(.128)	.509**(.212)	.411***(.160)	.238 (.285)
Black H	3.52**(1.79)	-1.32 (1.55)	1.84 (1.35)	-4.05**(1.87)	-4.66*(2.47)	-2.74 (3.37)	-4.87**(2.48)	-2.69 (3.44)
AmInd/AlaskNat	1.07***(.133)	.772***(.161)	.761***(.146)	.435**(.192)	-.280 (.651)	-.241 (.723)	-.468 (.657)	-.452 (.731)
Asian	1.07***(.301)	.180 (.400)	.325 (.317)	-.411 (.389)	-2.92***(.375)	-1.66***(.427)	-3.28***(.416)	-2.14***(.607)
Hawaiian/PI	9.31***(3.39)	8.22**(3.72)	7.81**(3.40)	6.81*(3.64)	-.807 (2.15)	-.098 (2.32)	-1.48 (2.17)	-.976 (2.05)
Share Civilian Labor Force	-.141*(.078)	-.231***(.076)	-.142*(.077)	-.214***(.075)	-.018 (.021)	-.022 (.026)	-.019 (.021)	-.022 (.027)
Average HH Size	-1.13***(.225)	-1.17***(.270)	-.762***(.233)	-.766***(.256)	-.207**(.085)	-.071 (.121)	-.209**(.085)	-.074 (.122)
Unemployment Rate	-.872***(.300)	-.546*(.324)	-.775***(.292)	-.428 (.313)	.094 (.079)	.087 (.115)	.081 (.080)	.075 (.118)
HH Inc X Average HH Size	.332***(.062)	.359***(.062)	.244***(.063)	.269***(.060)	.054*(.029)	.012 (.038)	.054*(.029)	.012 (.039)
(HH Inc X Average HH Size)^2	-.005***(.001)	-.005***(.001)	-.004***(.001)	-.004***(.001)	-.000 (.000)	-.000 (.001)	-.000 (.000)	-.000 (.001)
Share Poverty	.029***(.004)	.019***(.004)	.028***(.004)	.017***(.005)	.001 (.001)	.000 (.001)	.001 (.001)	.000 (.001)

Table 5 Cont'd: Determinants of non-filing rates of those owing at least \$500: IV-GMM regressions [1]

Variable	(1) Cross Section Sparse	(2) Cross Section Full	(3) Cross Section+Het Sparse	(4) Cross Section+Het Full	(5) Panel Sparse	(6) Panel Full	(7) Panel+Het Sparse	(8) Panel+Het Full
Education < High School[2]		.052 (.219)		-.063 (.219)		-1.07***(.346)		-1.09***(.345)
Some College[2]		.459* (.236)		.390* (.236)		-.123 (.429)		-.138 (.426)
Bachelors[2]		-.030 (.327)		-.073 (.320)		-1.16**(.530)		-1.18** (.545)
Masters/PhD[2]		.048 (.331)		-.042 (.323)		-1.00 (.755)		-1.10 (.754)
Share Inc.Self-employment[2]		2.10***(.345)		1.86***(.341)		1.09***(.370)		1.02***(.385)
Interest/Div/Rent[2]		-.068 (.338)		-.138 (.344)		.013 (.361)		-.058 (.336)
Social Security[2]		.560 (.841)		.203 (.876)		.910 (1.78)		.949 (1.78)
Supplemental SS[2]		-2.58 (3.41)		-1.79 (3.32)		-10.6* (5.47)		-10.8** (5.46)
Public Assistance[2]		17.1***(5.95)		16.6***(5.70)		28.2***(8.91)		28.3***(8.48)
Retirement Funds[2]		-.096 (.353)		-.091 (.354)		-1.74 (1.10)		-1.64 (1.12)
Other Sources[2]		1.84 (1.31)		2.12* (1.29)		5.46***(1.75)		5.49***(1.73)
Share Foreign Born[2]		1.13***(.291)		.942***(.286)		.515 (1.06)		.530 (1.05)
Speak Spanish at Home[2]		-.880* (.515)		.019 (.539)		.968 (.915)		.845 (1.06)
Speak Other Lang at Home[2]		-.171 (.145)		-.393 (.244)		-1.34** (.689)		-1.41 (1.34)
Share Married[2]		-1.63***(.362)		-1.83***(.402)		-.366 (.509)		-.348 (.529)
Share Wid/Div/Separated[2]		.574 (.720)		.230 (.754)		-.500 (.609)		-.490 (.631)
Mainline Protestant		-.040 (.115)		.035 (.129)				
Evangelical Protestant		-.120***(.044)		-.129***(.046)				
Catholic		-.086 (.066)		-.071 (.074)				
Orthodox		.450 (1.22)		.989 (1.23)				
Other Religion		-.012 (.124)		-.092 (.122)				
Ave. State IncomeTax Rate[2]						1.67 (1.83)		1.67 (1.83)
Med Yrs Res/Same Yr Ago[2]		-.033***(.005)		-.033***(.005)		-.923***(.357)		-.921** (.361)
Median Home Value[2]		.007***(.002)		.007***(.002)		-.003***(.001)		-.003** (.001)
Share HH Owner Occup.[2]		.460***(.166)		.431***(.165)		.697* (.379)		.737** (.376)
Heterogeneity by: HH Inc.[2]				.174 (.290)				.266 (.385)
Language[2]				.189 (.170)				.081 (.703)
Race			.397***(.078)	.404***(.086)			.275** (.140)	.333 (.290)
Religion				-.176** (.083)				
Centered R²	.712	.786	.723	.791	.017	.050	.018	.050
N	2276	2275	2276	2275	16467	16467	16467	16467

[1] All estimations are run on Stata 12.0 using ivreg2 and two step GMM. Numbers in parentheses are robust standard errors. *, **, and *** refer to two tailed significance at the 10, 5, and 1 percent levels, respectively. State dummies are included, but not reported. Enforcement Index Non-filers and Audit Rate Filers are instrumented.
[2] For panel models (5) – (8), these variables are available only at the state, rather than the county, level.

Appendix A: Variables, Definitions, and Sources

Variable:	Definition and Source
IRS VARIABLES:	Source: All IRS-provided variables are aggregated at the county level from internal IRS data and censored if there are fewer than 10 observations per county-year.
NON-FILING RATES	The broad or narrow non-filing rates are calculated by taking the number of non-filers (defined below) and dividing by the number of individuals who are identified as either filers or non-filers. The denominator is calculated as the sum of the number of non-filers, the number of filed returns by single and married-filing-separately taxpayers, and two times the number of filed returns by married-filing-jointly taxpayers.
FILERS	The total number of filed individual income tax returns for each county for the tax years 2000-2006. For purposes of creating the non-filing rates described above, counts were provided separately by filing status.
BROAD NON-FILING RATE NARROW NON-FILING RT 1 NARROW NON-FILING RT 2	The broad and both narrow definitions of non-filers are based on the set of individuals that IRS identified as not appearing on a tax return for the given tax year despite having filed a return in previous years and/or having third-party-provided information documents that indicate a filing requirement. "Broad non-filers" exclude deceased individuals from the set of IRS-identified non-filers. The first "Narrow non-filers" definition also excludes individuals who were late-filers, and is further limited to those who the IRS expected to have a positive balance due based on Information Returns Processing. The second "Narrow non-filers" definition is further limited to those whom the IRS expected to have a balance owing in excess of $500. County-level data for tax years 1999-2006.
ENFORCEMENT INDEX NON-FILERS	The ratio of the number of Delinquent Returns in the IRS Enforcement Revenue Information System (ERIS) where the first assessment was made against the taxpayer's account in the previous fiscal year over the (broad or narrow) number of IRS-identified non-filers for the prior tax year. The first assessment for the delinquent return is the first strong signal to the taxpayer that the IRS is seeking to recover a specific amount of unpaid taxes as well as penalties and interest. County-level data for 1999-2005.
AUDIT RATE FILERS	The ratio of the number of individual income tax audits that began in the previous calendar year over the number of filed returns in the prior tax year. County-level data for 1999-2005.

Appendix A (Cont'd): Definition of Variables

Variable:	Definition and Source
PENALTY RT NON-FILERS	Net penalties and interest assessed as a fraction of tax liability for individual taxpayers with delinquent returns, averaged over all such identified non-filers in the county . County-level data on a fiscal year basis for 2000-2006.
PENALTY RATE FILERS	Net penalties and interest assessed as a fraction of tax liability for individual taxpayers with filed returns, averaged over all such filers in the county. County-level data on a fiscal year basis for 2000-2006.
NN ENFORCEMENT INDEX NON-FILERS	The enforcement index against identified non-filers as defined above, for the nearest neighboring county whose population centroid is closest to the population centroid of a given county as of the 2000 census. Used as an instrument. Source: http://gothos.info/resource_files/CenPop2000_Mean_CO.txt
NN AUDIT RATE FILERS	The audit rate for filers as defined above, for the nearest neighboring county whose population centroid is closest to the population centroid of a given county as of the 2000 census. Used as an instrument. Source: http://gothos.info/resource_files/CenPop2000_Mean_CO.txt
PCT EITC	The percent of filed returns claiming eligibility for the Earned Income Tax Credit averaged over the two tax years prior to the tax year of interest. County-level data on a tax year basis for tax years 2000-2006. Used as an instrument.
INFO DOCS RATE	The number of third-party information documents sent to any address in a county, divided by the number of filers. This is averaged over the two tax years preceding the tax year of interest. County-level data on a tax year basis for tax years 2000-2006. Used as an instrument.
INFO DOCS TYPE	The number of *types* of third party information documents sent to any address in a county of the 23 types possible. This is averaged over the two tax years preceding the tax year of interest. County-level data on a tax year basis for tax years 2000-2006. Used as an instrument.

HETEROGENEITY MEASURES:

HH INCOME GINI	GINI coefficient of inequality of household income. Source: 2000-2005 annual data at state level constructed from American Community Survey data by Hisnanick and Rogers (undated). 2006 at state level from the American Community Survey, B19083. 2000 data at county level from the United States Census, customized data kindly provided by the US Census Bureau.

Appendix A (Cont'd): Definition of Variables

Variable:	Definition and Source

HETEROGENEITY MEASURES: (Cont'd)

RACE

Fragmentation measure based on the seven shares s_i of the population of identifying exclusively as white non-Hispanic, white Hispanic, black non-Hispanic, black Hispanic, Native American, Asian, or Hawaiian/ Pacific Islander. Calculated 2000 – 2006 annually at county level as

$$1 - \sum_{i=1}^{7} s_i^2.$$

LANGUAGE AT HOME

Fragmentation measure for shares s_i of population over 5 speaking English, Spanish or Other Language at home. $1 - \sum_{i=1}^{3} s_i^2$

Calculated at state level annually for 2000-2006, and county level for 2000.

RELIGION

Fragmentation measure based on the six shares s_i of adherents of Mainline Protestant, Evangelical, Catholic, Orthodox, Other, and Unclaimed for each county in 2000. Calculated as $1 - \sum_{i=1}^{6} s_i^2.$

SHARES UNDERLYING HETEROGENEITY MEASURES:

Race/Ethnicity:

WHITE NON-HISPANIC
WHITE HISPANIC
BLACK NON-HISPANIC
BLACK HISPANIC
AMER IND/ALASKAN NAT
ASIAN
HAWAIIAN/PACIFIC IS

Annual estimated share of county identifying as white non-Hispanic, white Hispanic, black or African American non Hispanic, black or African-American Hispanic, Native American or Alaskan, Asian, or Hawaiian or Pacific Islander. Estimated share identifying as "2 or more races" excluded.

Source: 1999-2006 annual data at county level from the Area Resource File (ARF) Access System 2009-2010. US Dept. of Health and Human Services, Health Resources and Services Administration, Bureau of Health Professions, Rockville, MD.

Appendix A (Cont'd): Definition of Variables

Variable:	Definition and Source

SHARES UNDERLYING HETEROGENEITY MEASURES: (Cont'd)

Language:

SPEAK ENGLISH AT HOME   Share speaking English, Spanish or Other Language at home, from the
SPEAK SPANISH AT HOME   population 5 and over.
SPEAK OTHER LANG HOME

Source: 2000-2006 annual data at state level from the American
Community Survey. 2000 data at county level from the United States
Census. Constructed 2000-2003 P034, 2004-2006 B06007

HOUSEHOLD INCOME   Real median household income deflated using the national CPI-U with
1999=100.

Source: 2000-2006 annual data at county level from the Area Resource
File (ARF) Access System 2009-2010. US Dept. of Health and Human
Services, Health Resources and Services Administration, Bureau of
Health Professions, Rockville, MD.

Religion Adherents: Share of each county in 2000 estimated to be adherents at mainline
MAINLINE PROTESTANT   Protestant churches, evangelical churches, a Roman Catholic church,
EVANGELICAL   an Orthodox church, another religious body, or to be unclaimed as an
CATHOLIC   adherent (the residual). In part because adherents of reporting churches
ORTHODOX   may live across county lines, 39 counties report adherents exceeding
OTHER RELIGION   their populations, which presumably understates adherents in
UNCLAIMED   adjacent counties. Rather than omit these counties, we redistribute their
negative UNCLAIMED shares to their other categories in proportion to
their normalized adherents shares. Thus shares sum to 1 in all counties.

Source: 2000 county data on unadjusted adherents comes from the
Association of Religion Data Archives (ARDA), collected by the
Association of Statisticians of American Religious Bodies. Members of
African American denominations and other religious bodies are thought
to be undercounted, and the "unclaimed" adherents are thought to
overstate the number of non-adherents.

OTHER COVARIATES:

POVERTY RATE   Percentage of individuals defined to be in poverty.

Source: 1999-2006 annual data at county level from the Area Resource
File (ARF) Access System 2009-2010. US Dept. of Health and Human
Services, Health Resources and Services Administration, Bureau of
Health Professions, Rockville, MD.

Appendix A (Cont'd): Definition of Variables

Variable:	Definition and Source
UNEMPLOYMENT RATE	Percentage of labour force aged 16+ who are unemployed.
	Source: 1999-2006 annual data at county level from the Area Resource File (ARF) Access System 2009-2010. US Dept. of Health and Human Services, Health Resources and Services Administration, Bureau of Health Professions, Rockville, MD.
SHARE MALE	Share of normally resident population male.
	Source: 1999-2006 annual data at county level from the Area Resource File (ARF) Access System 2009-2010. US Dept. of Health and Human Services, Health Resources and Services Administration, Bureau of Health Professions, Rockville, MD.

Age Distribution:
  SHARE AGE < 15
  SHARE AGE 15-19
  SHARE AGE 20-24
  SHARE AGE 25-44
  SHARE AGE 45-64
  SHARE 65 & OVER

Share of normally resident population aged under 15, 15-19, 20-24, 25-44, 45-64 or 65 & Over. Males and females summed.

Source: 1999-2006 annual data at county level, from the Area Resource File (ARF) Access System 2009-2010. US Dept. of Health and Human Services, Health Resources and Services Administration, Bureau of Health Professions, Rockville, MD.

SHARE FOREIGN BORN	Foreign born not of American parents/Total Population
	Source: 2000-2006 annual data at state level from the American Community Survey. 2000 data at county level from the United States Census. Constructed 2000-2003 P038, 2004-2006 B05002

Highest Education:
  EDUC < HIGH SCHOOL
  EDUC HIGH SCHOOL DIP
  EDUC SOME COLLEGE
  EDUC BACHELOR'S
  EDUC MAST/PH.D/PROF

Sum of male and female in each category of highest educational attainment divided by total population. "Some College" includes associate degrees.

Source: 2000-2006 annual data at state level from the American Community Survey. 2000 data at county level from the United States Census. Constructed 2000-2003 PCT034, 2004-2006 B15002

POPULATION DENSITY	Population per square mile annually at county level, defined as county population estimate divided by square miles in 2000.
	Source: 1999-2006 Area Resource File (ARF) Access System 2009-2010. US Dept. of Health and Human Services, Health Resources and Services Administration, Bureau of Health Professions, Rockville, MD.

Appendix A (Cont'd): Definition of Variables

Variable:	Definition and Source
AVE HOUSEHOLD SIZE	Average household size. Constructed as the difference between the estimated total population and population in group quarters, divided by the estimated number of housing units. Source: 1999-2006 annual data at county level from the Area Resource File (ARF) Access System 2009-2010. US Dept. of Health and Human Services, Health Resources and Services Administration, Bureau of Health Professions, Rockville, MD.
MEDIAN HOME VALUE	Real median value of owner occupied housing units. Nominal values deflated by annual average CPI-U (1999 = 100). Source: 2000-2006 annual data at state level from the American Community Survey. Constructed 2000-2003 H074, 2004-2006 B25077 2000 data at county level, undeflated, from the Area Resource File (ARF) Access System 2009-2010. US Dept. of Health and Human Services, Health Resources and Services Administration, Bureau of Health Professions, Rockville, MD.
HOME OWNERSHIP RATE	Share of occupied housing units that are owner occupied rather than renter occupied. Source: 2000-2006 annual data at state level from the American Community Survey. 2000 data at county level from the United States Census. Constructed 2000-2003 H035, 2004-2006 B25038
SAME RES YR AGO DIFF RES US YR AGO DIFF RES NOT US YR AGO	Percentage of the population 1 year and over who resided 1) in the same residence one year ago, 2) in a different residence in the United States one year ago, or 3) in a different residence outside the United States one year ago. Source: 2000-2006 annual data at state level from the American Community Survey. Constructed 2000-2003 P041, 2004-2006 B07202
MEDIAN YEAR IN RES	Median number of years householder has resided in current residence. Source: 2000-2006 annual data at state level from the American Community Survey. 2000 data at county level from the United States Census. Constructed 2000-2003 H36, 2004-2006 B25039

Appendix A (Cont'd): Definition of Variables

Variable:	Definition and Source
SH CIV POP IN LAB FORCE	Share of population aged 15-64 in the civilian labour force. Estimated as population in the civilian labour force aged 16+, divided by share of the estimated total population aged 15-64.
	Source: 1999-2006 annual data at county level from the Area Resource File (ARF) Access System 2009-2010. US Dept. of Health and Human Services, Health Resources and Services Administration, Bureau of Health Professions, Rockville, MD.
STATE INC TAX RATE	Average state level individual income tax rate, defined as total state individual income tax divided by total state personal income.
	Source: 2000-2006 fiscal year state level data on total state personal income from the Bureau of Economic Analysis http://www.bea.gov/regional/spi/drill.cfm. 2000-2006 fiscal year state level data on total state individual income tax kindly provided as customized order from the U.S. Census Bureau, Annual Survey of State and Local Government Finances and Census of Governments 1999-2006.

Marital Status:

NEVER MARRIED NOW MARRIED WID/DIV/SEPARATED	Percentage of the population 15 years and over whose current marital status is 1) never married, 2) currently married and spouse not absent because of separation, or 3) widowed/divorced/separated.
	Source: 2000-2006 annual data at state level from the American Community Survey. 2000 data at county level from the United States Census. Constructed 2000-2003 P031, 2004-2006 B12001

Share of Aggregate Household Income By Source:

WAGES AND SALARIES SELF-EMPLOYMENT INT/DIV/NET RENTAL SOCIAL SECURITY SUPPLEMENTAL SOC SEC PUBLIC ASSISTANCE RETIREMENT FUNDS OTHER	Percentage of households reporting income from 1) wage or salary, 2) self-employment, 3) interest, dividends or net rental, 4) social security, 5) supplementary social security, 6) public assistance, 7) retirement income sources, or 8) other sources. Or, percentage of aggregate household income in a region from these sources.
	Source:2000-2006 annual data at state level from the American Community Survey. 2000 data at county level from the United States Census. Constructed 2000-2003 P076, 2004-2006 B19052

Appendix B: Determinants of non-filing rates: linear cross section regression for Tax Year 2000[1]

Variable	(1) Broad NF Sparse	(2) Broad NF Full	(3) Broad NF+Het Sparse	(4) Broad NF+Het Full	(5) Narrow NF Sparse	(6) Narrow NF Full	(7) Narrow NF+Het Sparse	(8) Narrow NF+Het Full
Intercept	-1.00 (.634)	-1.79***(.618)	-2.32 (.649)	-2.28***(.642)	-1.37**(.625)	-2.07***(.624)	-2.89***(.621)	-2.93***(.643)
Penalty Rate Non-filers	-.000 (.001)	-.001 (.001)	-.000 (.001)	-.001 (.001)	-.001 (.001)	-.001 (.001)	-.000 (.001)	-.001 (.001)
Enforcement Index Non-filers	-.167***(.023)	-.065***(.021)	-.157***(.022)	-.071***(.020)	-.111***(.014)	-.055***(.012)	-.104***(.014)	-.056***(.012)
Penalty Rate Filers	.003***(.001)	.001 (.001)	.003**(.001)	.000 (.001)	.002 (.001)	-.000 (.001)	.002 (.001)	-.001 (.001)
Audit Rate Filers	4.81**(1.91)	4.77***(1.61)	5.43***(1.98)	5.34***(1.70)	3.92***(1.29)	3.85***(1.25)	4.64***(1.24)	4.42***(1.22)
Share Male	-.434 (.356)	.726**(.332)	-.861**(.345)	.543 (.332)	-.630*(.330)	.433 (.331)	-1.13***(.327)	.288 (.342)
Median HH Income	-.252**(.103)	-.314***(.098)	-.064 (.109)	-.211**(.100)	-.401***(.082)	-.409***(.073)	-.184**(.086)	-.271***(.076)
Age Shares 15-19	-5.97***(1.07)	-3.73***(.914)	-5.50***(.978)	-3.67***(.909)	-5.18***(1.06)	-3.40***(.939)	-4.64***(.955)	-3.29***(.938)
20-24	.822*(.438)	-.738 (.572)	1.55***(.439)	-.576 (.590)	.665 (.441)	-.791 (.561)	1.50***(.439)	-.585 (.584)
25-44	-1.37**(.683)	-1.57***(.541)	-.627 (.654)	-1.10**(.528)	-.802 (.652)	-1.25**(.541)	.054 (.626)	-.690 (.534)
25-64	1.13**(.504)	1.07**(.465)	1.95***(.499)	1.53***(.454)	1.14**(.494)	1.20***(.460)	2.09***(.482)	1.69***(.447)
65 Plus	-2.51***(.522)	-2.71***(.509)	-1.85***(.500)	-1.93***(.496)	-1.63***(.516)	-2.35***(.515)	-.872*(.494)	-1.54***(.504)
Share White H	.502***(.081)	.727**(.314)	.227**(.090)	-.084 (.374)	.485***(.075)	1.03***(.294)	.168**(.079)	.023 (.336)
Black NH	.901***(.051)	.692***(.063)	.548***(.074)	.316***(.093)	.813***(.050)	.597***(.064)	.406***(.075)	.209**(.099)
Black H	6.22***(2.07)	.261 (1.15)	4.35***(1.45)	-1.94 (1.36)	3.51**(1.52)	-1.32 (1.24)	1.35 (.931)	-3.96**(1.57)
AmInd/AlaskNat	1.21***(.102)	.800***(.111)	.857***(.121)	.440***(.133)	1.15***(.095)	.744***(.113)	.742***(.108)	.377***(.137)
Asian	.618**(.274)	-.748**(.307)	-.169 (.263)	-1.29***(.312)	.730***(.268)	-.427 (.314)	-.179 (.255)	-.983***(.312)
Hawaiian/PI	9.79***(2.60)	5.18**(2.32)	7.57***(2.60)	3.42 (2.36)	10.1***(2.59)	6.07***(2.29)	7.56***(2.52)	4.32*(2.36)
Share in Civ. Labor Force	-.193***(.060)	-.184***(.057)	-.189***(.060)	-.177***(.057)	-.206***(.060)	-.209***(.056)	-.202***(.060)	-.194***(.056)
Avg. HH Size	-1.03***(.184)	-.747***(.168)	-.595***(.198)	-.455***(.177)	-1.16***(.171)	-.859***(.153)	-.659***(.177)	-.499***(.158)
Unemployment Rate	-.624**(.249)	-.302 (.271)	-.486**(.246)	-.178 (.267)	-.640***(.229)	-.397*(.237)	-.480**(.219)	-.264 (.231)
HH Inc X Ave. HH Size	.268***(.054)	.246***(.049)	.164***(.056)	.178***(.051)	.327***(.045)	.295***(.040)	.206***(.046)	.213***(.041)
(HH Inc X HH Size)^2	-.005***(.001)	-.004***(.001)	-.004***(.001)	-.003***(.001)	-.005***(.001)	-.004***(.001)	-.004***(.001)	-.003***(.001)
Share in Poverty	.026***(.003)	.013***(.003)	.024***(.003)	.011***(.003)	.030***(.003)	.018***(.003)	.028***(.003)	.015***(.004)

Appendix B Cont'd: Determinants of non-filing rates: linear cross section regression for Tax Year 2000[1]

Variable	(1) Broad NF Sparse	(2) Broad NF Full	(3) Broad NF+Het Sparse	(4) Broad NF+Het Full	(5) Narrow NF Sparse	(6) Narrow NF Full	(7) Narrow NF+Het Sparse	(8) Narrow NF+Het Full
Education < High School		.179 (.160)		.101 (.162)		.289* (.156)		.155 (.154)
Some College		.821*** (.147)		.791*** (.147)		.764*** (.150)		.711*** (.149)
Bachelors		-.200 (.299)		-.219 (.283)		.280 (.208)		.170 (.211)
Masters/PhD		.983** (.473)		.947** (.478)		.417* (.252)		.329 (.247)
Share Inc. Self-employment		1.32*** (.209)		1.16*** (.209)		1.27*** (.209)		1.08*** (.210)
Interest/Div/Rent		-.014 (.232)		.052 (.231)		-.081 (.229)		-.153 (.233)
Social Security		1.79*** (.595)		1.24** (.615)		2.71*** (.569)		2.45*** (.578)
Supplemental SS		-3.92** (1.77)		-2.71 (1.80)		-5.65*** (1.77)		-4.47** (1.79)
Public Assistance		8.46*** (3.25)		8.91*** (3.38)		10.3*** (3.40)		11.3*** (3.50)
Retirement Funds		-.041 (.251)		-.017 (.256)		.029 (.241)		.019 (.243)
Other Sources		1.60* (.922)		1.67 (.926)		1.46* (.842)		1.74** (.822)
Share Foreign Born		1.32*** (.224)		1.17*** (.216)		1.25*** (.242)		1.07*** (.236)
Speak Spanish at Home		-.606 (.338)		.060 (.369)		-.994*** (.318)		-.156 (.332)
Speak Other Lang at Home		.008 (.118)		.000 (.194)		-.072 (.115)		-.184 (.182)
Share Married		-1.25*** (.229)		-1.44*** (.292)		-1.38*** (.271)		-1.57*** (.299)
Share Wid/Div/Separated		1.99*** (.436)		1.63*** (.471)		1.67*** (.432)		1.28*** (.472)
Mainline Protestant		-.211*** (.058)		-.160 (.061)		-.177*** (.056)		-.124** (.059)
Evangelical Protestant		-.057* (.034)		-.072** (.035)		-.028 (.032)		-.046 (.033)
Catholic		-.189*** (.040)		-.187*** (.044)		-.137*** (.038)		-.143*** (.041)
Orthodox		.628 (.887)		.826 (.954)		.979 (.966)		1.20 (1.04)
Other Religion		-.089 (.077)		-.126 (.080)		-.060 (.079)		-.110 (.081)
Median Years Residence		-.030*** (.004)		-.030*** (.004)		-.026*** (.004)		-.027*** (.004)
Median Home Value		.009*** (.002)		.009*** (.002)		.007*** (.002)		.007*** (.002)
Share HH Owner Occup.		.226* (.130)		.198 (.131)		.212* (.123)		.189 (.123)
Heterogeneity by: HH Inc.				-.117 (.236)				.252 (.202)
Language				.010 (.126)				.111 (.121)
Race			.434*** (.062)	.397*** (.066)			.501*** (.060)	.406*** (.069)
Religion				-.107** (.050)				-.099** (.048)
R²	.798	.849	.803	.853	.791	.838	.798	.843
N	2573	2572	2573	2572	2573	2572	2573	2572

[1] All estimations are run on Stata 12.0 using ivreg2 and two step GMM. Numbers in parentheses are robust standard errors. *, **, and *** refer to two tailed significance at the 10, 5, and 1 percent levels, respectively. State dummies are included, but not reported.

Appendix B Cont'd: Determinants of non-filing rates: linear fixed effects regression for Tax Years 2000-2006[1]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	Broad NF Sparse	Broad NF Full	Broad NF+Het Sparse	Broad NF+Het Full	Narrow NF Sparse	Narrow NF Full	Narrow NF+Het Sparse	Narrow NF+Het Full
Intercept	-3.72***(.384)	-2.30**(.889)	-3.69***(.384)	-2.36**(.922)	-4.09***(.275)	-2.90***(.713)	-4.07***(.276)	-2.94***(.768)
Penalty Rate Non-filers	-.000 (.000)	-.000 (.000)	-.000 (.000)	-.000 (.000)	-.000 (.000)	-.000 (.000)	-.000 (.000)	-.000 (.000)
Enforcement Index Non-filers	-.014** (.007)	-.009 (.007)	-.014**(.007)	-.008 (.007)	-.007* (.004)	-.003 (.006)	-.007* (.004)	-.003 (.006)
Penalty Rate Filers	.000 (.000)	-.000 (.001)	.000 (.000)	.000 (.001)	.000 (.000)	.000 (.001)	.000 (.000)	.000 (.001)
Audit Rate Filers	1.61***(.278)	2.16***(.612)	1.67***(.281)	2.19***(.617)	2.32***(.269)	1.89***(.376)	2.35***(.272)	1.91***(.392)
Share Male	-.863***(.263)	-.746**(.330)	-.872***(.271)	-.753**(.341)	-.627**(.292)	-.686**(.339)	-.631**(.296)	-.689**(.344)
Real Median HH Income	-.027 (.095)	.028 (.097)	-.027 (.095)	.023 (.097)	-.071 (.051)	.010 (.063)	-.071 (.051)	.007 (.063)
Age Shares 15-19	.227 (.469)	.118 (.767)	.118 (.471)	-.038 (.787)	.288 (.413)	.233 (.667)	.234 (.417)	.139 (.668)
20-24	1.05***(.221)	.848* (.458)	.964***(.221)	.784* (.452)	.583***(.209)	.720* (.399)	.540***(.210)	.674* (.397)
25-44	1.18***(.231)	.993* (.497)	1.08***(.233)	.899* (.496)	.690***(.228)	.719* (.382)	.642***(.230)	.657* (.381)
45-64	.890***(.259)	.872 (.566)	.826***(.259)	.833 (.558)	.175 (.241)	.447 (.418)	.143 (.241)	.420 (.413)
65 Plus	1.02***(.341)	1.24* (.710)	.985***(.341)	1.21* (.708)	1.34***(.330)	1.39**(.673)	1.32***(.330)	1.36**(.675)
Share White H	1.55*** (.205)	1.21***(.304)	1.10*** (.253)	.753***(.278)	1.82***(.218)	1.71***(.403)	1.59*** (.248)	1.41***(.364)
Black NH	.916*** (.189)	.844***(.287)	.523** (.205)	.406 (.283)	.621***(.182)	.555** (.237)	.427** (.201)	.279 (.291)
Black H	-9.32** (3.66)	-7.70 (5.43)	-9.86***(3.66)	-7.73 (5.15)	-.515 (2.71)	-.343 (3.30)	-.782 (2.72)	-.427 (3.38)
AmInd/AlaskNat	.733 (.652)	.492 (.790)	.481 (.654)	.270 (.793)	1.46** (.630)	1.25**(.562)	1.33 (.643)	1.10* (.601)
Asian	.524 (.514)	-.347 (1.01)	-.102 (.527)	-1.00 (.911)	-.201 (.400)	.025 (.685)	-.509 (.448)	-.402 (.658)
Hawaiian/PI	.824 (1.53)	1.55 (2.06)	-.476 (1.53)	-.157 (1.80)	.705 (1.59)	1.81 (2.42)	.064 (1.62)	.769 (2.08)
Share in Civ. Labor Force	.022 (.021)	.022 (.032)	.020 (.021)	.022 (.031)	-.007 (.022)	-.017 (.029)	-.008 (.022)	-.017 (.029)
Avg. HH Size	-.067 (.144)	.045 (.150)	-.071 (.144)	.038 (.149)	-.094 (.093)	.016 (.123)	-.096 (.093)	.013 (.123)
Unemployment Rate	.460***(.091)	.587***(.156)	.438***(.092)	.567***(.158)	.400***(.082)	.408** (.172)	.389***(.083)	.395**(.173)
HH Inc X Ave. HH Size	.033 (.054)	.005 (.055)	.033 (.054)	.008 (.055)	.066** (.030)	.029 (.042)	.066** (.030)	.030 (.041)
(HH Inc X HH Size)^2	-.002* (.001)	-.001 (.001)	-.001* (.001)	-.001 (.001)	-.002***(.000)	-.002***(.001)	-.002***(.000)	-.002***(.001)
Share in Poverty	-.001* (.001)	-.001 (.001)	-.001* (.001)	-.001 (.001)	.000 (.001)	.000 (.001)	.000 (.001)	-.000 (.001)

Appendix B Cont'd: Determinants of non-filing rates: linear fixed effects regression for Tax Years 2000-2006[1]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	Broad NF Sparse	Broad NF Full	Broad NF+Het Sparse	Broad NF+Het Full	Narrow NF Sparse	Narrow NF Full	Narrow NF+Het Sparse	Narrow NF+Het Full
Education < High School[2]		-.787 (.569)		-.850 (.579)		-1.26***(.465)		-1.29***(.474)
Some College[2]		-.689 (.496)		-.777 (.513)		-.542 (.493)		-.588 (.504)
Bachelors[2]		-.478 (.914)		-.553 (.902)		-.676 (.598)		-.715 (.618)
Masters/PhD[2]		-.128 (.831)		-.493 (.866)		-.458 (.872)		-.649 (.895)
Share Inc. Self-employment[2]		.899* (.489)		.874* (.496)		.816** (.374)		.800** (.377)
Interest/Div/Rent[2]		-.770 (.636)		-.826 (.593)		-.554 (.702)		-.589 (.660)
Social Security[2]		-3.52** (1.67)		-3.55**(1.65)		-1.68 (1.78)		-1.68 (1.76)
Supplemental SS[2]		-6.72 (8.95)		-7.02 (8.98)		2.29 (7.06)		2.12 (7.19)
Public Assistance[2]		56.9** (23.8)		55.8** (23.0)		29.4** (12.0)		28.8** (11.4)
Retirement Funds[2]		-.711 (1.45)		-.549 (1.46)		-1.26 (1.39)		-1.17 (1.41)
Other Sources[2]		3.22 (2.18)		3.20 (2.15)		4.69**(1.99)		4.69** (1.97)
Share Foreign Born[2]		-3.03* (1.70)		-3.05* (1.64)		1.13 (.877)		1.12 (.857)
Speak Spanish at Home[2]		3.17***(1.04)		1.64 (1.49)		.903 (.815)		.175 (1.25)
Speak Other Lan at Home[2]		-.284 (1.41)		-2.06 (2.24)		-1.38 (.919)		-2.22 (1.79)
Share Married[2]		-1.19* (.608)		-1.24* (.628)		-1.02* (.560)		-1.05*(.571)
Share Wid/Div/Separated[2]		-1.18 (.908)		-1.21 (.944)		-.442 (.757)		-.459 (.776)
Avg. State Income Tax Rate[2]		4.28* (2.49)		4.28* (2.50)		1.94 (2.32)		1.94 (2.30)
Share Same House Yr Ago[2]		-1.18** (.491)		-1.23** (.489)		-1.11** (.500)		-1.13** (.511)
Real Median Home Value[2]		.002 (.002)		.003 (.002)		-.002 (.001)		-.001 (.001)
Share HH Owner Occup.[2]		1.03* (.587)		1.15* (.590)		.815 (.527)		.877 (.529)
Heterogeneity by: HH Inc.[2]				.245 (.458)				.143 (.410)
Language[2]				1.12 (1.06)				.534 (.850)
Race			.504***(.168)	.555***(.186)			.249 (.166)	.350 (.227)
Within R^{2}	.725	.741	.725	.742	.659	.670	.659	.670
Between R^{2}	.333	.204	.430	.282	.232	.204	.269	.248
Overall R^{2}	.371	.254	.455	.327	.270	.241	.306	.283
N	18436	18436	18436	18436	18436	18436	18436	18436

[1] All estimations are run on Stata 12.0. Numbers in parentheses are robust standard errors. Standard errors clustered to the state level in models (2), (4), (6) and (8). *, **, and *** refer to two tailed significance at the 10, 5 and 1 percent levels, respectively. Year dummies are included but not reported.
[2] These variables are available only at the state, rather than the county, level.