DEPARTMENT OF ECONOMICS AND FINANCE

COLLEGE OF BUSINESS AND ECONOMICS

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

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

James Alm, Jeremy Clark and Kara Leibel

WORKING PAPER

Department of Economics and Finance
College of Business and Economics
University of Canterbury
Private Bag 4800, Christchurch
New Zealand

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

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

July 24th, 2011

Abstract:

In this paper we present a rare empirical study on the determinants of tax filing compliance in the United States using county and state level data from 2000 to 2006. As well as including explanatory variables identified in the rational compliance framework such as audit and penalty rates, we examine the role of social capital on tax compliance. In particular, we test whether county level heterogeneity in income, language, race and religion can explain variation in filing rates. While several issues are not yet addressed in this preliminary analysis (such as censoring, endogeneity between audit and non-filing rates, and people's self-selection to county), our preliminary findings are that non-filing rates are falling in the enforcement rate, rising in the penalty rate for income reporting noncompliance, and rising in home-ownership and unemployment rates. Regarding heterogeneity, non-filing rates do not seem to covary with household income inequality or home language fragmentation, but increase with racial fragmentation.

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. Please address all correspondence to this author.

[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 steady rise in studies identifying a negative correlation, and possibly causation, between measures of social or income 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, be members of organizations, contribute to fundraisers, trust others, or support welfare programs (Alesina and La Ferrara 2000, 2002; Costa and Kahn 2003a, 2003b; Putnam 2007; Miguel and Gugerty 2005; Clark and Kim 2011; Leigh 2006; Gustavsson and Jordahl 2008; Luttmer 2001). 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 heterogeneity 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 voluntary indicator of social capital, tax filing compliance behavior. Second, we provide a rare empirical study on the determinants of actual tax filing compliance, using data derived from individual filing decisions.

Perhaps for reasons of data availability, most studies that have looked at heterogeneity's effects on social capital have concentrated on 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 the broader compulsory area of tax compliance has been examined in a much more limited fashion, owing to the very limited amount of compliance data available to researchers, especially data that is based in some way on *individual* decisions. 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 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. These aggregate studies notwithstanding, we are not aware of any paper that has examined the effects of heterogeneity on *actual* tax compliance, based on *individual* filing decisions, again because such data are simply not widely available.

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

---

[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."

measure compliance with the United States' federal tax-filing requirements using previously unreleased Internal Revenue Service (IRS) data on identified non-filers. In recent years the IRS has identified approximately nine million cases annually of those individuals who have either stopped filing without clear cause, or for whom third-party income reports indicate filing should be expected.[2] With access to these data, we are able to estimate known non-filing rates for almost all 3140 counties of the United States, using both a broad measure of the proportion of identified non- or late filers, as well as a narrower measure of the proportion of identified non-filers whom the IRS believes have a high likelihood of still owing a positive amount to the federal government. We can then examine the effect of heterogeneity at either the county or the state level on filing non-compliance rates, while controlling for other factors shown in the tax compliance literature to predict income tax non-compliance rates (Dubin, Graetz, and Wilde 1990; Beron, Tauchen, and Witte 1992; Erard and Ho 2001; Gentry and Kahn 2009). Based on data availability, we examine heterogeneity by race at the county level for all 7 years; we examine heterogeneity by language and household income at the county level for 2000 and the state level for all 7 years, and we examine heterogeneity by religion at the county level for 2000 only.

The second contribution of our study is to provide a more general investigation of the determinants of non-compliance with tax filing requirements in the United States. The absence of compliance data has often limited researchers in their ability to test the various theoretical models of compliance using empirical data. Even when these data have been available, they have typically been for *reporting* non-compliance, using information on individuals who file tax returns but do not declare all taxable income (Clotfelter 1983; Beron, Tauchen, and Witte 1992;

---

[2] We say "identified" non-filers, because there likely exist other United States residents who have never filed a tax return (when required by law to do so), and for whom the IRS receives no third party (e.g. bank or employer) income reports. These are the "unknown non-filers", who cannot be addressed by our study.

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 who are required by law to file tax returns in fact file these returns. Again, lack of data on non-filing "ghosts" has largely been responsible for this omission. One notable exception is a study by Erard and Ho (2001), who use 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 compliance as well as on income reporting compliance, using audit data from the Annual Report of the Commissioner of Internal Revenue, 1977-1986. They find that audit rates are positively associated with reporting compliance, but negatively associated with filing rates (possibly because they 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 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 as well as heterogeneity on reporting and filing compliance. We then present our data and empirical framework in section 3. (An Appendix gives details of the data construction.) In section 4 we present our estimation results. Section 5 concludes.

2. Theoretical Framework

Most theoretical analysis of the individual's tax compliance decision focuses on his or her *reporting* decision, assuming implicitly that the individual has already chosen to file a return. This basic theoretical model begins with the economics-of-crime model of Becker (1968), first applied to tax compliance by Allingham and Sandmo (1972). Here a rational individual is viewed as maximizing 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 and productive 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.[3]

More precisely, an individual who has already decided to file a tax return and report income 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).*

---

[3] See Cowell (1990), Andreoni, Erard, and Feinstein (1998), Alm (1999), Slemrod and Yitzhaki (2002), Sandmo (2005), and Alm (2011) 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 $(1-p)$, or $(1-p) U(I-tR-C)$; the second if the individual is caught with probability p and

is forced to pay all taxes on unreported income, or $p U(I-tR-(1+f)t(I-R)-C)$.[4] Optimization of

equation (1) 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. Indeed, the traditional portfolio analysis of the *reporting* decision does not fully

capture the individual's *filing* decision because submitting a tax return with underreported

liabilities is inherently different from failing to submit a return at all. 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 audit 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.[5]

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).

Here, an individual who decides to file a tax return and report income R has expected utility

---

[4] In this framework, we ignore the distinction that is sometimes made between the probability of audit and the probability of detection of underreported income; here we assume these are identical.
[5] In practice, in the US, non-filers are generally pursued by IRS through means other than what is technically considered an "audit," but for purposes of this discussion, we use the term "audit" loosely to capture the IRS enforcement with respect to non-filers.

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

Here p^* is the probability that an individual with income I who has not filed a return is

apprehended by an audit, 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. A standard feature of nearly all 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). The presence of 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; when $R≥I_{TPR}$, then p will be much lower and decreasing rapidly. The presence of third-information also affects the probability of audit for 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} 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 will 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 noncompliance becomes pervasive, then the social norm of compliance disappears.[6] Also, to the extent that heterogeneity affects social capital in the form of the frequency of people's interactions with others, it may affect social norms regarding compliance, or alternatively 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.[7] 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 social heterogeneity affects social capital, it would then affect γ, either by changing social norms, or the importance individuals assign to abiding by them. 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 experiences no change in utility from the psychic cost of evasion.

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

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

and similarly the filing decision becomes:

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

---

[6] 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.
[7] 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 must first choose whether or not to file a tax return, by comparing the value of expected utility from non-filing in equation (2)' with the expected utility of filing and reporting the optimal amount of income R in equation (1)'.

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 and the penalty rates), the comparative statics results are more straightforward, indicating that an increase in, say, the audit rate for non-filing decreases the payoff to non-filing. The impact of heterogeneity via social capital and social norms is not clear-cut, because it would change the psychic cost of cheating in both the reporting (equation (1)') and filing decisions (equation (2)').[8]

Regardless, however, it is straightforward to demonstrate in this framework that the likelihood of non-filing depends upon an array of fiscal factors, like the probabilities of audit for filers and 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 audits and of detection of non-filing. Non-filing will also depend upon socio-demographic variables that proxy for unobserved preferences or costs associated with filing (e.g., marital status, education or age). Finally, of relevance to our work here, heterogeneity may affect social capital, and thus either alter social norms regarding compliance, or the weight people place on conforming to those norms. Heterogeneity can be captured by dispersion measures of salient

---

[8] 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.

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.

3. Data and Empirical Specification

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, rates of enforcement initiated on identified non-filed returns, and effective penalty rates on delinquent returns as well as all other tax returns. To these we have added census and other non-IRS variables used in the tax compliance literature to measure income, source of income as it affects third party reporting and probability of detection, pecuniary benefits from filing such as the Earned Income Tax Credit, and demographic characteristics to proxy for unobserved preferences regarding compliance, as well as social heterogeneity measures. All variables are described in the Appendix, and summary statistics are provided in Tables 1 and 2. Here we briefly summarize the key or unusual 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 (ty), and divide it by the total number of individual filers plus non-filers.[9] We construct both a "broad" measure of non-filing that includes late-filers, and a "narrow" measure that is restricted to those who do not file and are thought to have positive tax owing. Our broad measure thus attempts to capture *all* (identifiable)

---

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

11

non-filers at the cost of including some who may not have needed to file. Our narrow measure attempts to avoid including the ex post "innocent", at the cost of excluding some of the ex post "guilty."

Our measures of audit and penalty rates work as follows. We define the "audit rate" of filed income tax returns as the number of individual income tax audits initiated in a calendar year divided by the number of returns filed in that year (reflecting income earned in the previous year).[10] We assume that this audit rate affects the filing decision and therefore the rate of non-filers in the subsequent calendar year. We define an "enforcement rate" for (identified) non-filers as the number of individuals who receive a "delinquent return" notice on their account in a given calendar year divided by the number of (broad or narrow) individuals who were identified as not filing a return in that year (for income earned in the previous tax year).[11] Again, we assume that this non-filer enforcement rate will influence the filing decision made in the subsequent year.

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.[12] Despite the stability of the penalty formula over time and across counties, there may still be slight variation in effective penalty rates due to abatements, time to detection and thus interest charged, and similar considerations. We thus define a non-filing penalty rate as the ratio of net

[10] This is a ratio rather than a "true" audit rate, in which the audits in the numerator would be a subset of the audits in the denominator and would be based on the tax year of the return (the year during which the income was earned). The timing becomes difficult and perhaps less salient for taxpayer decision-making because audits can be initiated up to several years after the filing of the return. Instead, we define our audit rate as a ratio that captures audit activity during a single calendar year.

[11] The establishment of a "delinquent return" on a taxpayer's account signals the beginning of the enforcement process for non-filers, which begins with notices to the taxpayer requesting them to file and, if the taxpayer does not comply, eventually leads to a Substitute For Return, where the IRS prepares a return and assesses tax liability based on third-party-reported income information.

[12] The interest rate charged is the federal short-term rate plus 3 percent and compounds daily.

penalties plus interest charged on delinquent returns, divided by net tax assessed for delinquent returns; this is calculated on a fiscal year basis for taxpayers for whom the most recent enforcement revenue activity for the delinquent return was in that fiscal year.[13] To provide a proxy penalty rate for filers, we use the ratio of net penalties plus interest charged to taxpayers for all *but* delinquent returns, divided by net tax assessed to all taxpayers *but* delinquent returns. Again, this is on a fiscal year basis according to the timing of the most recent enforcement revenue activity. These effective penalty rates are assumed to affect the filing decisions made in the following calendar year.

It is important to note that all IRS-generated variables (regarding number of non-filers, audits, enforcements, or penalties assessed) are censored for counties with fewer than 10 observations. For 2000, for example, of 3140 counties, filing rates are available for 96 percent. However non-filing enforcement rates are available for only 74 percent, audit filing rates for 84 percent, and penalty rates (for non filing or incorrectly filed returns) for 98 percent.[14] With censoring unaddressed in analysis, those counties that are sparsest in population and/or most law-abiding may be absent.

Other variables thought to influence the filing choice, such as level of income, or probability of detection via source of income or occupation, or filing benefits such as eligibility for the Earned Income Tax Credit (EITC), are taken from the United States Census of 2000 at county level, the American Community Survey (ACS) at state level for 2000 to 2006, and the Bureau of Economic Analysis at state level, and the Area Resource File at county level. To proxy for eligibility for the EITC, which provides tax credits for low income working households

---

[13] These data come from IRS' Enforcement Revenue Information System, which tallies activity on a fiscal year basis.

[14] Non-filing enforcement rates are available for a higher share of counties in subsequent years, such that for the 2000-2006 panel overall, 83 percent of county-years are available.

that increase in household size, we have included the share of the population in the civilian labor force, the unemployment rate, average household size and median household income, and the interaction of the household size and income. 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 ACS. Religious adherence rates at county level are also included for 2000 only, sourced from the Association of Religion Data Archives (ARDA). Median (owner-occupied) home values are also included to proxy for differences in cost of living that affect the real cost of income tax. Finally, following Dubin et al. (1990), we include average state individual income tax rates. Given the information sharing that exists between state and federal tax agencies, individuals in high tax states may face higher incentives not to file federal returns.

Our last class of variables, heterogeneity measures, are defined as follows. For language, race and religion, we follow the social capital literature in constructing fragmentation measures,

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

That is, a fragmentation index divides a population into n categories. Each category's share of the population is squared, summed, and subtracted from 1. (It 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 available from the ARDA: Mainline Protestant, Evangelical

Protestant, Roman Catholic, Orthodox, Other Religions, and Unclaimed. Finally, we measure the dispersion of household income F_{Income}, using the Gini coefficient as provided by the census and ACS, which like the fragmentation index is bounded between 0 and 1 and increasing in heterogeneity.

Of our four measures of heterogeneity, the data pertaining to race are the most detailed, as population counts (census) or estimates (ARF) are available annually at county level for 2000-2006. Our measures of heterogeneity by language spoken at home and household income are available at county level for 2000 only, but at state level for 2000-2006. Finally, our measure of heterogeneity by religion is available at county level for 2000 only.

3.2 Empirical Estimation Strategy

To exploit the greater number of explanatory variables available at county level for the calendar/tax year 2000, we proceed in two steps. First, we use a cross section regression of the form:

(4) $\qquad y_{ij,2000} = X'_{ij,2000}\beta + u_{ij,2000}$,

where $y_{ij,2000}$ refers to the (broad or narrow) non-filing rate in county i of state j, and $X_{ij,2000}$ is a vector of county level characteristics and state dummies, and $u_{ij,2000}$ is a random error. With limited correlations between our four heterogeneity measures, (none exceeding $|.54|$), we follow this with a specification that adds all of them simultaneously, or:

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

where $F_{ij,2000}$ is the vector of F_{Income}, $F_{Language}$, F_{Race}, and $F_{Religion}$ for county i in state j.

We then compare these results with those from a fixed effects panel regression that follows counties between (calendar/tax years) 2000 and 2006. Here some explanatory variables X are available at county level ij, whereas others Z are available only at state level j:

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

Here we also include year dummies, and cluster standard errors at state level. As before, with limited correlations between our heterogeneity measures, (non exceeding |.42|), we follow this with a specification that adds all of them simultaneously:

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

4. Results

Descriptive statistics for all variables are provided in Tables 1 (cross section) and 2 (panel). We begin by examining simple correlations between each heterogeneity measure and non-filing rates. Figure 1 provides scatterplots of year 2000 non-filing rates, both broadly and narrowly defined, for each measure at the county level. No strong correlations appear between income, language or religious fragmentation and non-filing rates. Of the four dimensions, a positive correlation appears only between racial fragmentation and both broad and narrow non-filing rates.

[Figure 1 about here.]

More formally, for income inequality, the pairwise correlation is .35 and .42 for broad and narrow non-filing rates, respectively. For language fragmentation, the correlation is .39 and .30, respectively, while for religious fragmentation it is -.23 and -.24, respectively. Reducing to the three dimensions of heterogeneity available for 2000-2006 non-filing rates, broad and narrow correlations are weaker for (state level defined) income inequality at .21 and .22, respectively. Correlations are also weaker for (state level defined) language fragmentation, at .24 and .18, respectively. However correlations are higher for (county level defined) racial fragmentation, at .67 and .68, respectively.

16

Similar pairwise correlations can be examined between our enforcement variables and non-filing rates. Since these are available at county level for all years, we consider only the pooled 2000-2006 data. For the penalty rate for non-filers, the correlation with broad and narrow non-filing rates is -.13 and -.14, respectively. For the enforcement rate for non-filers, the correlation is -.15 and -.21, respectively. In contrast, the audit rate for filers is positively correlated, at .39 for broad non-filing rates, and .44 for narrow non-filing rates. This might suggest, as Dubin et al. (1990) found, that a more severe audit regime for filers can have the side effect of raising non-filing rates. Finally, the correlation between the penalty rate for filing taxpayers and the non-filing rate is low and of an unexpected sign, at -.07 (broad non-filing) and -.08 (narrow non-filing).

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

Beginning with year 2000 cross section regression, results are provided in Table 3. Without fragmentation measures included, we find that the broad non-filing rate in column (1) is rising in filed audit rates as in the simply correlation. Of course, interpreting the role of IRS audit rates on non-filing rates is complicated by possible endogeneity; the IRS is likely to devote more enforcement resources where noncompliance is suspected. Subsequent drafts of this paper will address this possibility using several potential instruments. Surprisingly, we find that broad non-filing is also rising in the enforcement rate against detected non-filers. The latter could reflect problems with our measure of non-filer enforcement rate for the earliest years of our sample (tax years 1999 to 2002), as the internal IRS data source from which we are drawing may not have complete taxpayer account activity that far back. (We address this provisionally in the

following panel regression with year dummies.) Compared to those whose highest education is high school graduation, broad non-filing rates are rising in the share with some college or associate degree, but falling in the share with a bachelor's degree. Regarding age, broad non-filing rates are increasing in the share of a county aged 45-64, and lowest for those with a large share of people under 20 or over 65. Consistent with the issue of third party reporting, broad non-filing rates are increasing in the share of household income from self-employment (which must be self-reported to the IRS), rather than wages and salaries.[15] Relative to the share unmarried, broad non-filing rates are falling in the share currently married, but rising in the share widowed, divorced, or separated.   Interestingly, the non-filing rate is not affected by the share who speak Spanish or Other languages rather than English at home when the share foreign born is controlled for, but the share foreign born raises non-filing rates. Non-filing rates are falling in residential stability, i.e. the median number of years individuals have lived at the same residence, but rising in median home value. The latter is consistent with nominal taxes having a higher real cost in areas with a higher cost of living.

[Table 3 about here.]

The non-filing rate does not vary significantly by home ownership rates, mean household size, or unemployment rate. In particular, factors germane to eligibility for the EITC, such as interaction between household size and income, are not significant.

---

[15] For this and all following specifications we also try replacing the 8 income source share variables with the 13 industry shares used by the census and ACS. These are: agriculture, construction, manufacturing, wholesale, retail, transportation, information, finance/insurance, professional, education/health/social, arts/entertainment, public administration, and other. In every specification, relative to the share of the civilian labor force in agriculture, non-filing rates are significantly different at the 5 percent of 10 percent level only in the information sector, where they are lower.

When we shift our focus to explaining variations in the *narrow* non-filing rate in column (3), our results are very similar, with the following exceptions. The troublesome enforcement rate for non-filers loses significance but retains the "wrong" sign. Those with a bachelor's degree are now no more likely to file than high school graduates. Counties with a higher share of income from Social Security relative to wages and salary are now less likely to file, while non-filing is now decreasing in the share speaking Spanish or another language at home (with the share foreign born controlled). Finally, consistent with EITC eligibility, narrow non-filing rates are now decreasing in the share of the working age population in the civilian labor force.

Finally, when we add our four heterogeneity measures to either the broad (column 2) or narrow (column 4) non-filing specifications, the estimated effects of our non-heterogeneity measures remain remarkably stable. With this as background, the effect of heterogeneity on non-filing rates is only modest. Broad non-filing rates (including later filers and those who, upon later inspection, owe no further money) is actually falling in household income inequality and religious fragmentation, but each is significant only at the 10 percent level. Narrow non-filing rates are also falling in religious fragmentation, but no longer varying in income inequality, and are now rising in language fragmentation.

Thus, based on cross section evidence from the 2000 census year, the effects of social heterogeneity on non-filing rates are significant in only 4 of 8 cases; in 3 of these 4 cases measured heterogeneity is actually *increasing* filing rates. Perhaps heterogeneity is good for social capital in this instance?

However, it is well known that in cross section regressions unmeasured influences on non-filing rates can cause misleading correlations between included variables. To better control

for those influences on non-filing that are unobserved but time invariant, we move next to fixed effects results.   All results are provided in Table 4.

[Table 4 about here.]

Beginning as before with broad non-filing rates and heterogeneity excluded, we find in column (1) that non-filing is not affected by (slight) variations in the effective penalty rate as in cross section. However non-filing is now falling in the enforcement rate for detected non-filers, in line with expectations. Non-filing is now marginally rising in the effective penalty rate for filers, but surprisingly no longer affected by the audit rate of filed returns. Unlike in cross section, broad non-filing is now falling in the share male, and no longer affected by education. The effects of the age distribution is also attenuated; relative to the share of a county aged 45-64, only a rise in the share 15-19 lowers the non-filing rate.

Perhaps most surprisingly, broad non-filing rates are no longer affected by the share of a county's household income that comes from self-employment rather than (third-party verified) wages and salaries. This finding was robust to alternative measures of income source, such as the share of households reporting income from self-employment, or the share of the civilian labour force self-employed in private, for profit incorporated or unincorporated businesses. Instead, relative to the share of income from wages and salaries, broad non-filing rates differed only in the share coming from Social Security (lower). Non-filing rates are also no longer affected by marital status shares. With share foreign born controlled, non-filing is now positively associated with the share speaking Spanish at home, but only at the 10 percent level, while the share foreign born itself is now marginally negatively associated with non-filing.

In other effects, broad non-filing no longer varies significantly with real median home value. Home ownership rates are now co-vary positively with non-filing, while more of the

20

variables associated with EITC eligibility are now significant, including average household size, unemployment, and the interaction between income and household size. The average state income tax rate, which could not be tested in cross section, is positively signed but insignificant. Thus we do not find significant evidence that residents of high income tax states avoid filing federally to avoid combined state and federal taxes.

When we again shift our focus to explaining variations in the *narrow* non-filing rate over time in column (3), our results are again generally similar, with the following exceptions. Narrow non-filing rates are no longer lower for share male, suggesting perhaps that female filers are only more likely to either file late, or not file when no income is still owed. Also of interest, narrow non-filing rates are rising in real median household income, unlike broad non-filing rates. Similarly, while broad non-filing rates do not vary in the share of income derived from public assistance, narrow non-filing rates increase markedly. Narrow non-filing is again falling in the share married, suggesting as with gender that married couples may simply be more likely to file late or neglect to file when no money is owed. When the share foreign born is controlled, narrow non-filing is no longer increasing in the share speaking Spanish at home, and is falling in the share speaking another non-English language. Narrow non-filing is also no longer affected by residential stability, nor by the share foreign born. Conversely, narrow non-filing is falling in real median home value. All other effects are qualitatively similar to those for broad non-filing.

Let us turn finally to addressing the effects of heterogeneity. When we add our three available heterogeneity measures to either the broad (column 2) or narrow (column 4) non-filing specifications in Table 4, the estimated effects of our non-heterogeneity measures remain stable, with the exception of some of the coefficients on the language and race shares. These are the shares that underlie our language and racial fragmentation measures, and thus correlation

21

between these shares and their respective dispersion measures may account for this. In particular, with fragmentation measures added, the share speaking Spanish at home loses its positive association with broad non-filing.   With this as background, we find that heterogeneity no longer dampens some non-filing rates as in cross section. Instead, for both broad and narrow non-filing rates, income inequality and fragmentation by home language lose any effect. In contrast, racial fragmentation shifts from having no effect in cross section, to having a definite positive association with non-filing rates under fixed effects.

5. Conclusions

This paper has attempted to make two contributions. First, it has joined Dubin et al. (1990) and Erard and Ho (2001) in providing a rare empirical study of the determinants of tax filing compliance in the United States. Using individual United States tax data aggregated to county level for the tax years 2000 to 2006, we have found that tax non-filing rates have varied in ways that consistent with augmented versions of the rational compliance model. In particular, fixed effects analysis suggests that non-filing rates are decreasing in the enforcement rate taken against detected non-filers, and increasing in the effective penalty rate applied to filed but misreported taxes. Non-filing rates are also decreasing in the factors that proxy for eligibility for the Earned Income Tax Credit, such as household size interacted with median household income and the employment rate.   Non-filing rates are also increasing in home ownership rate, and narrow non-filing in particular is increasing in real median household income.

Among demographic factors, non-filing rates are decreasing in the population share of counties below the age of 20, or in the share of a county's household income from Social Security.   While non-filing does not vary by educational achievement shares, narrow non-filing

rates in particular are decreasing in the share married, in real median home value, and in the share who speak a language other than English or Spanish at home. While broad non-filing rates are negatively associated with residential stability, narrow non-filing rates are less conclusively related. In contrast to the findings of Dubin et al. (1990) and Erard and Ho (2001), identified non-filing rates did not vary with the share of income from self-employment. While this finding may seem counter-intuitive because of the lack of third party verification of self-employment income, we would point out that the availability (lack) of third party income reports may increase (decrease) both the likelihood that non-filing will be detected, and that filed reported income will be audited. Thus it is not clear *ex ante* what the effect of self-employed income should be on non-filing rates.

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 according to household income, language spoken at home, race, or (for 2000 only) religion has any effect upon non-filing rates. While cross section analysis suggests that income and religious heterogeneity might actually lower non-filing rates, these effects were modest, and disappear in fixed effects regression that better accounts for unobservable influences on non-filing that are constant over time. Instead, under fixed effects regression, both broadly- and narrowly- defined non-filing rates do not vary with heterogeneity by income or language. But they do both increase significantly with racial fragmentation, even when the underlying racial shares of a county are controlled for.

With regards to social capital, our estimated effects of heterogeneity under panel regression (neutral or negative) are less cheerful than those in cross section (neutral or positive). Yet they are consistent with the findings of previous studies estimating the effects of racial

heterogeneity on compulsory activities such as census return rates (Vigdor 2004), or 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 2010). On the other hand, it should be emphasized that various dimensions of heterogeneity are *not* being found to reduce social capital indicators, and some may in fact increase them.

Nonetheless, given that the United States and indeed all OECD countries are growing ever more racially diverse (Putnam 2007), it might be fruitful to explore ways to reduce or reverse the effects of this dimension of heterogeneity on both compulsory and voluntary indicators of social capital. As discussed in Li (2010), tax agencies that emphasize common benefits from a well-functioning tax system might reduce filing noncompliance. More broadly, in increasingly diverse societies, governments who are concerned with social capital may wish to set policies that emphasize shared citizenship, values, or identity.

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 (2011). Measuring, explaining, and controlling tax evasion: Lessons from theory, experiments, and field studies. *International Tax and Public Finance*, forthcoming.

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.

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 (2011). The effect of neighbourhood diversity on volunteering: Evidence from New Zealand. Working Paper, University of Canterbury. Christchurch, New Zealand.

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 (1-2), 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.

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 (1) 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.

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: Scatter Plots of County Level Non-filing Rates by Fragmentation in the Year 2000 (N = 3024)
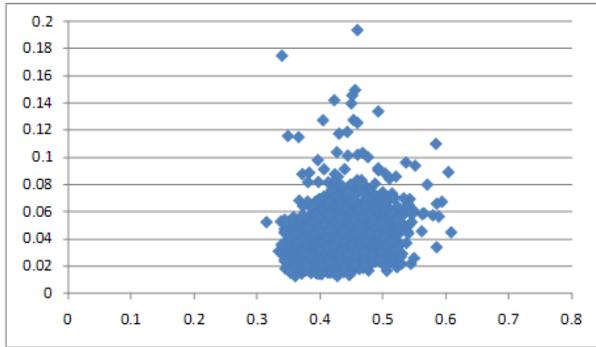
A. Broad Non-filing Rate as a Function of
    Household Income Inequality (Gini)
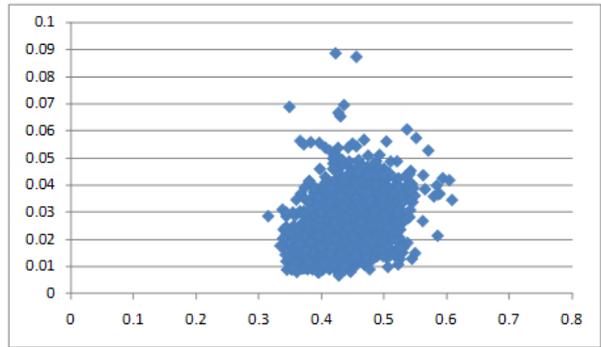
B. Narrow Non-filing Rate as a Function of
    Household Income Inequality (Gini)
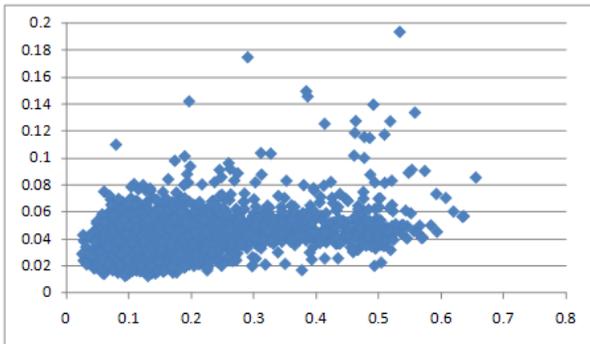
C. Broad Non-filing Rate as a Function of
    Fragmentation by Home Language
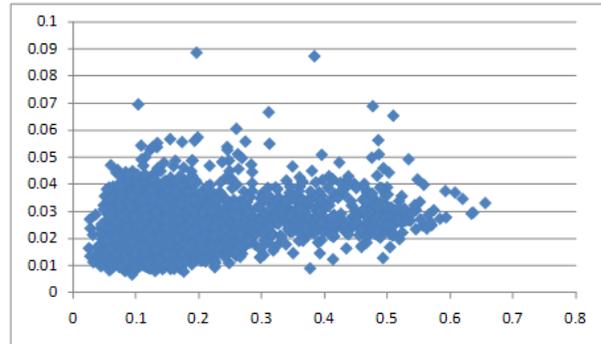
D. Narrow Non-filing Rate as a Function of
    Fragmentation by Home Language

**Figure 1 (Cont'd):** Scatter Plots of County Level Non-filing Rates by Fragmentation in the Year 2000
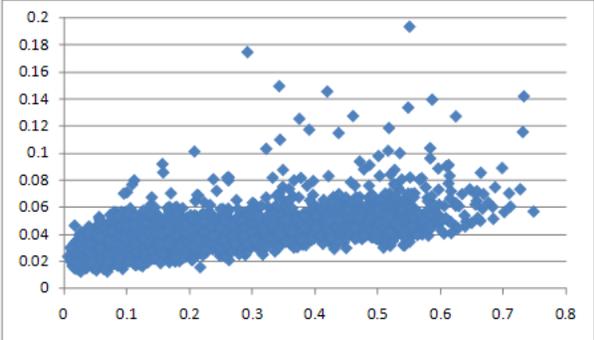
E. Broad Non-filing Rate as a Function of
    Fragmentation by Race
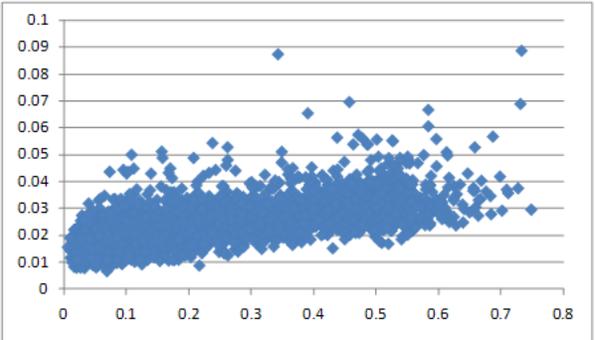
F. Narrow Non-filing Rate as a Function of
    Fragmentation by Race
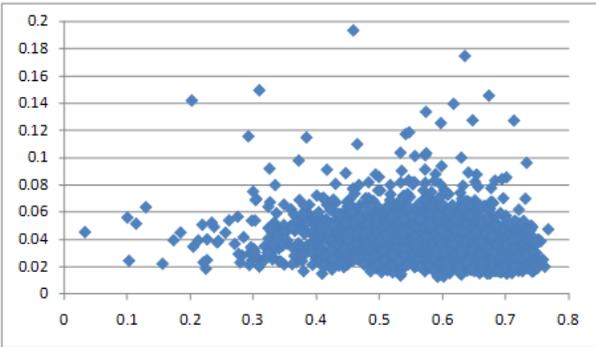
G. Broad Non-filing Rate as a Function of
    Fragmentation by Religion
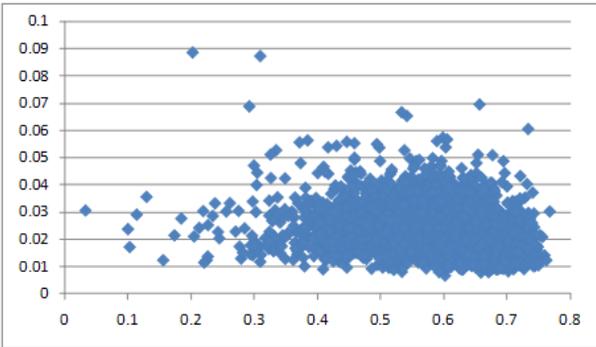
H. Narrow Non-filing Rate as a Function of
    Fragmentation by Religion

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

Variable	N	Mean	St.Dev.	Min	Max
Broad Non-filing Rate	3026	.0391	.0154	.0122	.1938
Narrow Non-filing Rate	3026	.0236	.0088	.0068	.0886
IRS enforcement variables not presented					
Share Male	3140	.4962	.0222	0	.6758
Real Median Household Inc.	3138	36372	9005	15231	91210
Educ. < High School	3140	.2262	.0876	.0304	.6530
Educ. High School Diploma	3140	.3469	.0655	.1093	.5325
Educ. Some College/Assoc Deg	3140	.2617	.0567	.0952	.4489
Educ. Bachelor's	3140	.1097	.0492	0	.4002
Educ. Masters/PhD/Prof Deg	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 Soc 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
Never Married	3140	.2248	.0562	.0902	.5613
Currently Married (not sep)	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 Lang. 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	3140	84579	47730	0	1000001

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

Variable	N	Mean	St.Dev.	Min	Max
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 Amer Ind/Alaskan Nat	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 (x 100)	3140	4.764	2.618	0	27.6
Share Civ Pop in Labor Force	3139	.7359	.1123	0	1.5633
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
Share of Civ Lab Force in					
Agric/Fishing/Forest/Mining	3140	.0723	.0763	0	.5819
Construction	3140	.0771	.0240	.0107	.2414
Manufacturing	3140	.1581	.0914	0	.4855
Wholesale	3140	.0298	.0114	0	.1332
Retail	3140	.1148	.0207	0	.2690
Transportation	3140	.0549	.0189	0	.2605
Information	3140	.0187	.0103	0	.1073
Finance	3140	.0455	.0187	0	.2071
Professional	3140	.0528	.0265	0	.2347
Education/Health/Social	3140	.2028	.0443	.0655	.4706
Arts/Entertainment	3140	.0712	.0332	0	.3645
Other	3140	.0478	.0095	0	.0975
Public Administration	3140	.0541	.0311	0	.4263

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

Variable	N	Mean	St.Dev.	Min	Max	State or County Level?
Broad Non-filing Rate	21132	.0354	.0143	.0066	.2051	County
Narrow Non-filing Rate	21127	.0208	.0086	.0041	.0886	County
IRS enforcement variables not presented						
Share Male	21982	.4966	.0208	0	.7010	County
Real Median Household Inc.	21987	33702	8658.0	0	88243	County
Educ. < High School	21987	.1650	.0431	.0858	.2590	State
Educ. High School Diploma	21987	.3109	.0399	.1932	.4269	State
Educ. Some College/Assoc Deg	21987	.2771	.0321	.1685	.3839	State
Educ. Bachelor's	21987	.1612	.0261	.0891	.2312	State
Educ. Masters/PhD/Prof Deg	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 Soc 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
Never Married	21987	.2648	.0254	.2136	.5402	State
Currently Married (not sep)	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 Lang. at Home	21987	.0474	.0293	.0087	.2486	State
Same Residence Year Ago	21987	.8372	.0222	.7617	.8949	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?
Share Foreign Born	21987	.0720	.0562	.0083	.2724	State
Real Median Home Value	21987	116471	47711	69679	442696	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 Amer Ind/Alaskan Nat	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
Unemployment Rate (x100)	21971	5.489	2.512	0	27.6	County
Share Civ 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
Share of Civ Lab Force in						
Agric/Fishing/Forest/Mining	21983	.0277	.0207	.0002	.1220	State
Construction	21983	.0743	.0111	.0272	.1216	State
Manufacturing	21983	.1317	.0417	.0102	.2416	State
Wholesale	21983	.0357	.0045	.0065	.0474	State
Retail	21983	.1191	.0065	.0507	.1538	State
Transportation	21983	.0531	.0079	.0304	.0952	State
Information	21983	.0250	.0063	.0132	.0725	State
Finance/Real Estate	21983	.0656	.0105	.0381	.1284	State
Professional	21983	.0855	.0193	.0500	.2105	State
Education/Health/Social	21983	.2048	.0192	.1158	.2617	State
Arts/Entertainment	21983	.0808	.0160	.0573	.2678	State
Other|	21983	.0479	.0041	.0329	.1015	State
Public Administration	21983	.0488	.0148	.0301	.1683	State

Table 3:
Determinants of non-filing rates: cross section regression for Census Year 2000 (N = 2,265 counties)[1]

Variable	(1) Broad Non-filing Rate		(2) Broad Non-filing Rate +Fragmentation		(3) Narrow Non-filing Rate		(4) Narrow Non-filing Rate + Fragmentation	
Intercept	.155***	(.034)	.172***	(.035)	.093***	(.016)	.088***	(.017)
Penalty Rate Nonfilers	-.003	(.003)	-.003	(.003)	-.002	(.002)	-.002	(.002)
Enforcement Rt Nonfilers	.071***	(.022)	.068***	(.023)	.009	(.006)	.009	(.006)
Penalty Rate Filers	.003	(.008)	.001	(.008)	.002	(.005)	.002	(.005)
Audit Rate Filers	.244**	(.107)	.256**	(.104)	.103***	(.035)	.106***	(.036)
Share Male	.025	(.020)	.021	(.020)	.007	(.011)	.005	(.011)
Median HH Income	-.000	(.000)	-.000	(.000)	-.000	(.000)	-.000	(.000)
Education < High School	-.010	(.009)	-.005	(.010)	-.002	(.004)	-.003	(.005)
Some College	.019***	(.007)	.019***	(.007)	.011***	(.004)	.009**	(.004)
Bachelors	-.046**	(.022)	-.038**	(.019)	-.005	(.006)	-.006	(.006)
Masters/PhD	.042	(.039)	.049	(.041)	-.007	(.007)	-.007	(.007)
Age Shares Under 15	-.056**	(.028)	-.061**	(.027)	-.037***	(.013)	-.041***	(.013)
15-19	-.199***	(.046)	-.207***	(.046)	-.113***	(.026)	-.117***	(.027)
20-24	-.100***	(.028)	-.097***	(.028)	-.053***	(.015)	-.052***	(.016)
25-44	-.075***	(.021)	-.077***	(.022)	-.044***	(.011)	-.042***	(.011)
65 Plus	-.155***	(.025)	-.144***	(.024)	-.092***	(.013)	-.086***	(.013)
Share of HH Income from								
Self-employment	.065***	(.012)	.062***	(.012)	.041***	(.007)	.037***	(.007)
Interest/Div/Rent	.028**	(.013)	.036***	(.013)	.011*	(.006)	.009	(.007)
Social Security	.038	(.032)	.009	(.034)	.058***	(.017)	.055***	(.017)
Supplemental SS	-.091	(.104)	-.024	(.110)	-.110*	(.061)	-.101	(.063)
Public Assistance	.136	(.277)	.219	(.304)	.228	(.164)	.271	(.179)
Retirement Funds	.014	(.014)	.015	(.015)	.011	(.007)	.011	(.007)
Other Sources	-.020	(.048)	-.029	(.051)	.004	(.025)	.010	(.025)
Married	-.073***	(.015)	-.072***	(.016)	-.046***	(.009)	-.046***	(.009)
Widowed/Divorced/Sep	.044*	(.026)	.048*	(.026)	.027*	(.015)	.026*	(.015)
Speak Spanish at Home	-.022	(.021)	-.010	(.025)	-.027***	(.010)	-.014	(.011)
Speak Other Lng at Home	-.004	(.008)	-.017	(.013)	-.007*	(.004)	-.018***	(.006)
Median Yrs at Residence	-.001***	(.000)	-.001***	(.000)	-.001***	(.000)	-.001***	(.000)

Table 3 (Cont'd):
Determinants of non-filing rates: cross section regression for Census Year 2000 (N = 2,265 counties)

	(1)	(2)	(3)	(4)
Variable	Broad Non-filing Rate	Broad Non-filing Rate +Fragmentation	Narrow Non-filing Rate	Narrow Non-filing Rate + Fragmentation
Share Foreign Born	.057*** (.012)	.056*** (.003)	.037*** (.007)	.035*** (.007)
Median Home Value	.000*** (.000)	.000*** (.000)	.000*** (.000)	.000*** (.000)
Home Ownership Rate	.002 (.008)	.003 (.009)	.004 (.004)	.005 (.004)
Ave HH Size	-.010 (.007)	-.010 (.007)	-.005 (.003)	-.002 (.003)
Unemployment Rate	-.000 (.000)	.000 (.000)	-.000 (.000)	-.000 (.000)
Share in civ. Lab Force	-.005 (.003)	-.005* (.003)	-.003** (.002)	-.003** (.002)
Median HH Income X AveHH Size	.000 (.000)	.000 (.000)	.000* (.000)	.000 (.000)
Fragmentation by:				
HH Income (Gini)		-.030* (.017)		.002 (.006)
Language		.011 (.008)		.009** (.004)
Race[2]		.001 (.004)		.001 (.002)
Religion[2]		-.005* (.003)		-.003** (.001)
R^2	.805	.807	.818	.819
N	2265	2265	2265	2265

[1] Run on Stata 10.0. Numbers in parentheses are robust standard errors. *, **, and *** refer to two tailed significance at the 10, 5 and 1 percent levels, respectively. State dummies included, but not reported.

[2] The underlying share variables for the race and religion fragmentation indexes were included as control variables. For religion, the share variables included Mainline Protestant, Evangelical, Catholic, Orthodox, and Other Religion, with non-adherents the omitted share. For race, the share variables included White Hispanic, Black NonHispanic, Black Hispanic, American Indian/Alaska Native, Asian, and Hawaiian/ Pacific Islander, with White NonHispanic the omitted group.

Table 4:

Determinants of non-filing rates: fixed effects panel regression for 2000-2006[1]

Variable	(1) Broad Non-filing Rate		(2) Broad Non-filing Rate +Fragmentation		(3) Narrow Non-filing Rate		(4) Narrow Non-filing Rate + Fragmentation	
Intercept	.094**	(.040)	.101**	(.040)	.033*	(.017)	.035**	(.016)
Penalty Rate Nonfilers	.000	(.000)	.000	(.000)	.000	(.000)	.000	(.000)
Enforcement Rt Nonfilers	-.008**	(.003)	-.008**	(.003)	-.003***	(.001)	-.003***	(.001)
Penalty Rate Filers	.001*	(.000)	.001*	(.000)	.000*	(.000)	.000**	(.000)
Audit Rate Filers	.045	(.035)	.049	(.034)	-.007	(.016)	-.004	(.015)
Share Male	-.039**	(.019)	-.041**	(.020)	-.010	(.009)	-.010	(.009)
Real Median HH Income	.000	(.000)	.000	(.000)	.000***	(.000)	.000***	(.000)
Education < High School[2]	-.022	(.023)	-.026	(.021)	-.012	(.009)	-.014	(.008)
Some College[2]	.012	(.022)	.003	(.022)	.010	(.011)	.007	(.011)
Bachelors[2]	-.012	(.063)	-.018	(.061)	-.016	(.015)	-.018	(.014)
Masters/PhD[2]	-.009	(.045)	-.015	(.045)	.010	(.017)	.004	(.016)
Age Shares Under 15	-.039	(.026)	-.034	(.018)	-.026**	(.011)	-.024**	(.011)
15-19	-.062**	(.026)	-.071**	(.027)	-.025*	(.014)	-.028*	(.015)
20-24	-.004	(.013)	-.004	(.014)	.004	(.007)	.003	(.007)
25-44	.003	(.015)	.001	(.015)	.002	(.008)	.007	(.008)
65 Plus	.009	(.019)	.011	(.016)	.016	(.011)	.017	(.011)
Share HH Income from[2]								
Self-employment	.012	(.027)	.016	(.026)	.003	(.008)	.005	(.010)
Interest/Div/Rent	-.006	(.024)	-.003	(.024)	.006	(.014)	.007	(.014)
Social Security	-.127*	(.064)	-.133**	(.065)	-.086**	(.034)	-.087**	(.033)
Supplemental SS	-.494	(.437)	-.500	(.436)	.203	(.147)	.202	(.149)
Public Assistance	2.76	(1.75)	2.65	(1.69)	.609**	(.263)	.585**	(.249)
Retirement Funds	-.047	(.050)	-.042	(.049)	-.014	(.024)	-.013	(.024)
Other Sources	.002	(.087)	-.003	(.089)	.043	(.036)	.043	(.037)
Married[2]	-.031	(.033)	-.037	(.034)	-.022**	(.011)	-.023**	(.012)
Widowed/Divorced/Sep[2]	-.063	(.042)	-.067	(.044)	-.009	(.015)	-.010	(.015)
Speak Spanish at Home[2]	.128*	(.068)	.019	(.063)	-.016	(.025)	-.030	(.035)
Speak Oth Lan at Home[2]	.014	(.077)	-.117	(.083)	-.037**	(.015)	-.054*	(.032)
Sh Same Res Year Ago[2]	-.046**	(.018)	-.051***	(.018)	-.015	(.009)	-.016*	(.010)

Table 4 (Cont'd):
Determinants of non-filing rates: fixed effects panel regression for 2000-2006[1]

Variable	(1) Broad Non-filing Rate	(2) Broad Non-filing Rate +Fragmentation	(3) Narrow Non-filing Rate	(4) Narrow Non-filing Rate + Fragmentation
Share Foreign Born[2]	-.202* (.120)	-.200* (.119)	-.012 (.025)	-.009 (.024)
Real Med Home Value[2]	.000 (.000)	.000 (.000)	-.000** (.000)	-.000** (.000)
Home Ownership Rate[2]	.050** (.013)	.056*** (.013)	.027*** (.009)	.029*** (.008)
Ave HH Size	.008* (.004)	.007* (.004)	.006*** (.002)	.006*** (.002)
Unemployment Rate	.000*** (.000)	.000*** (.000)	.000*** (.000)	.000*** (.000)
Share in civ. Lab Force	-.000 (.001)	.000 (.001)	-.000 (.001)	-.000 (.001)
Median HH Income X AveHH Size	-.000** (.000)	-.000* (.000)	-.000*** (.000)	-.000*** (.000)
Average State Inc Tax Rt	.180 (.112)	.177 (.112)	.025 (.044)	.025 (.043)
Fragmentation by:				
HH Income (Gini)[2]		-.004 (.017)		-.002 (.009)
Language[2]		.080 (.058)		.010 (.019)
Race[3]		.038*** (.012)		.020*** (.007)
R^2 within	.608	.611	.660	.661
R^2 between	.054	.117	.097	.159
R^2 overall	.097	.158	.135	.193
N	18341	18341	18341	18341

[1] Run on Stata 10.0. Numbers in parentheses are standard errors clustered at state level. *, **, and *** refer to two tailed significance at the 10, 5 and 1 percent levels, respectively. Year dummies included but not reported.
[2] Refers to variables available only at the state, rather than county, level.
[3] The underlying share variables for the race fragmentation index were included as control variables. Specifically, these included White Hispanic, Black NonHispanic, Black Hispanic, American Indian/Alaska Native, Asian, and Hawaiian/ Pacific Islander, with White NonHispanic the omitted group.

Appendix: 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 fewer than 10 observations per county-year.
NON-FILING RATE	Both the broad and 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. This 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 by tax year for 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	Both the broad and 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. "Narrow non-filers" further excludes individuals who were late-filers, and are further limited to those who the IRS expected to have a positive balance due based on Information Returns Processing. County-level data for 1999-2006.
ENFORCEMENT RATE NON-FILERS	The ratio of the number of Delinquent Returns established by IRS in the previous calendar year over the (broad or narrow) number of IRS-identified non-filers for the prior tax year. The establishment of the delinquent return indicates the beginning of the enforcement process for nonfilers. 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.
PENALTY RT NON-FILERS	Average net penalties and interest assessed as a fraction of tax liability for taxpayers with delinquent returns. County-level data on a fiscal year basis for 2000-2006.

Appendix (Cont'd): Definition of Variables

Variable:	Definition and Source
PENALTY RATE FILERS	Average net penalties and interest assessed as a fraction of tax liability for all other taxpayers (excluding those with delinquent returns). County-level data on a fiscal year basis for 2000-2006.

FRAGMENTATION 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.
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.
SPEAK ENGLISH V WELL	Fragmentation measure for shares s_i of population over 5 speaking English very well, speaking Spanish but not English very well, and speaking Other language but not English very well, $1-\sum_{i=1}^{3}s_i^2{}_L$ 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 $1-\sum_{i=1}^{6}s_i^2.$

Appendix (Cont'd): Definition of Variables

Variable:	Definition and Source

SHARES UNDERLYING FRAGMENTATION MEASURES:

Race/Ethnicity:

WHITE NON-HISPANIC
WHITE HISPANIC
BLACK NON-HISPANIC
BLACK HISPANIC
AMER IND/ALASKAN NAT
ASIAN

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.

HAWAIIAN/PACIFIC IS

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.

Language:

SPEAK ENGLISH AT HOME
SPEAK SPANISH AT HOME
SPEAK OTHER LANG HOME

Share speaking English, Spanish or Other Language at home, from the population 5 and over.

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

SP ENG VERY WELL
SP SPAN; ENG NOT V WELL
SP OTH; ENG NOT V WELL

Share of population 5 and over who speak English as native language *or* as second language but who speak it "Very Well." Share of population 5 and over who speak Spanish as native language, and who speak English less than "Very Well." Share of population 5 and over who speak Other language as native language, and who speak English less than "Very Well."

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.

Appendix (Cont'd): Definition of Variables

Variable:	Definition and Source

Religion Adherents:
MAINLINE PROTESTANT
EVANGELICAL
CATHOLIC
ORTHODOX
OTHER RELIGION
UNCLAIMED

Share of each county in 2000 estimated to be adherents at mainline Protestant churches, evangelical churches, a Roman Catholic church, an Orthodox church, another religious body, or to be unclaimed as an adherent (the residual). In part because adherents of reporting churches may live across county lines, 39 counties report adherents exceeding their populations, which presumably understates adherents in 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.

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.

Appendix (Cont'd): Definition of Variables

Variable:	Definition and Source

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

CITIZENSHIP RATE

Not a US citizen/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 P037, 2004-2006 B05001

Highest Education:

EDUC < HIGH SCHOOL
EDUC HIGH SCHOOL DIP
EDUC SOME COLLEGE
EDUC BACHELOR'S
EDUC MASTERS/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

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.

Appendix (Cont'd): Definition of Variables

Variable:	Definition and Source

Distribution of HH Size:

1 PERSON
2 PEOPLE
3 PEOPLE
4 PEOPLE
5 PEOPLE
6 OR MORE PEOPLE

Share of households in each county in 2000 with 1 person, 2 people, 3 people, 4 people, 5 people, or 6 or more people.

Source: 2000 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

MORTGAGE

Percentage of all housing units with a mortgage, contract to purchase, or similar debt.

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 H078, 2004-2006 B25081

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

Appendix (Cont'd): Definition of Variables

Variable:	Definition and Source
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
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.
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.
STATE/LOC TAX BURDEN	Annual state level per capita tax burden from in-state and out-of-state state and local taxes. (Divides total state and local taxes paid by average per capita income at state level). Source: 2000-2006 annual state level data calculated by The Tax Foundation http://www.taxfoundation.org/taxdata/show/335.html

Appendix (Cont'd): Definition of Variables

Variable:	Definition and Source

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

Industry:

AGRICULTURE
CONSTRUCTION
MANUFACTURING
WHOLESALE
RETAIL
TRANSPORTATION
INFORMATION
FINANCE/INSURANCE
PROFESSIONAL
EDUCATION/HEALTH
ARTS/ENTERTAINMENT
OTHER SERVICES
PUBLIC ADMINISTRATION

Percentage of civilian population 16 years and over employed in these industries. Shares sum to 100%.

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 P066, 2004-2006 B24030

Class of Worker:

EMP W/S PVT PR
EMP W/S PVT NON-PR
SELF-EMP PVT PR INC
SELF-EMP PVT PR NONINC
GOV'T WORKER
UNPAID FAM WORKER

Percentage of civilian population 16 years and over who are: 1) employed wage or salary workers in private, for profit firms, 2) employed wage or salary workers in private, non-profit firms, 3) self-employed workers in incorporated private, for-profit firms, 4) self-employed workers in non-incorporated private, for-profit firms, 5) employees of the federal, state or local government, and 6) unpaid family workers.

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 P068, 2004-2006 B24080

Appendix (Cont'd): Definition of Variables

Variable: Definition and Source

Marital Status:

NEVER MARRIED   Percentage of the population 15 years and over whose current marital
NOW MARRIED   status is 1) never married, 2) currently married and spouse not absent
WID/DIV/SEPARATED   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