IDENTITY AND TAX EVASION*

Zhiyong An, University of California at Berkeley

INTRODUCTION

In this research, we are relating tax evasion to taxpayers' identity (Akerlof and Kranton, 2000, 2002, and 2005). The basic idea is straightforward. According to Akerlof and Kranton, the concept of identity implies that if an individual's actual behavior deviates from the ideal behavior associated with the individual's identification, then the individual will lose utility. If we apply the concept of identity in the context of tax compliance, the intuition is clear: (1) People are identified with taxpaying; (2) The ideal behavior (norms) associated with this identification is that people think they should pay the tax; and (3) If people evade tax and thus their actual behavior departs from the ideal behavior, they will lose an extra utility. Under this interpretation, people would differ by how much they are identified with taxpaying.

A great deal of anecdotal evidence shows that identity plays a very significant role in people's tax compliance decision. Here, we quote some words we find from the Internet. The context of these quotes is the antitax movement in the United States. "For Holland patriotism meant continuing to avoid taxes… Holland's Freedom Council sold 'small business manuals' which consisted of advice on tax evasion… Holland did more than just advocate this belief; he lived it…" (http://www.militia-watchdog.org/holland.asp). Clearly, Holland is not identified with taxpaying at all so that he evades tax and even provides advice for other people to evade tax. "They have convinced themselves they don't have to pay taxes…There is no legal authority for the federal government to enforce the federal income tax on average Americans." (http://www.newsmax.com/archives/articles/2002/11/15/153033.shtml). It is also clear that these people (*We The People Congress, an antitax organization*) are also not identified with taxpaying so that they don't pay taxes.

In order to test this idea, we first build a simple theoretical model by introducing the concept of identity into the standard model on tax compliance (Allingham and Sandmo, 1972) and show that given an exogenous actual income, the more identified with taxpaying a taxpayer is, the more income he (or she) will declare to the tax authorities (i.e., there will be less tax evasion). Our identity-based utility function captures two important ideas not represented by utility functions of standard economics. First, it captures psychologists' and sociologists' view that decisions depend on social category. In our utility function, a taxpayer's utility varies with his category. Second, it captures the notion of norms and ideal type, since the taxpayer's utility depends on the deviation of his declared income from the ideal declaration that varies with his category. These two points are not surprising because in a model of utility, a person's identity describes "gains and losses in utility from behavior that conforms or departs from the norms for particular social categories in particular situations" (Akerlof and Kranton, 2005). This simple model is a variation of the standard model on tax compliance (Allingham and Sandmo, 1972) and thus also shows that identity theory can be perfectly compatible with standard economic theory.

Second, we have analyzed the 1987 Taxpayer Opinion Survey (TOS) data (Harris and Associates, Inc., 1988). The 1987 TOS is available from the Inter-university Consortium for Political and Social Research (ICPSR). This survey asks taxpayers' opinions and evaluations of the U.S. tax systems; especially, the respondents are asked about their attitudes toward tax evasion and tax authorities. In addition, demographic information such as age, gender, race, income, education, marital status, employment, etc. are collected in these surveys. The universe of these surveys is all adult taxpayers in the United States and the sampling method is stratified random sample. From the 1987 TOS data, we construct the proxy for individual identity (from Question 1a and Question 103b) and the proxy for individual tax evasion (Question 44). We can also extract demographic information such as age, gender, race, income, education, marital status, employment, etc. Then we regress the proxy for tax

* I thank George A. Akerlof and Alan J. Auerbach for valuable discussions and advice. Comments from Kim M. Bloomquist are highly appreciated. Financial assistance from the Burch Center for Tax Policy and Public Finance, University of California at Berkeley and the Institute of Business and Economic Research, University of California at Berkeley was gratefully acknowledged.

evasion on the proxy for identity and other control variables (demographic variables) and check the estimated coefficient of the proxy for identity. The basic result is that the estimated coefficient of the proxy for identity is negative and statistically significant, which is consistent with the identity theory: the more identified with taxpaying a taxpayer is, the less tax evasion there will be.

One criticism of the analysis of the 1987 TOS data is that it might really study the relationship between views on tax compliance and views of goods, not actual tax compliance. This concern drives us to go beyond survey data analysis and do a cross-state analysis. Like the literature on tax compliance taking a cross-section approach at aggregate level (Witte and Woodbury, 1985, Dubin and Wilde, 1988, Beron et al., 1992, Dubin et al., 1987 and 1990, Plumley, 1996), we employ the methodology of linear regression analysis. We use the state-level Wage and Salary Gap for year 2000 produced by the U.S. Bureau of Economic Analysis (BEA) as the measure of state-level tax evasion (Brown et al., 2006). We construct indexes for state-level identity from different data sources including the 1987 TOS, the voting and registration rates in the November 2000 U.S. presidential election, and the state-level murder rate in 2000. Then we do the cross-state analysis and check the estimated coefficient of the index for state-level identity. The basic result is that the estimated coefficient of the index for state-level identity is negative and statistically significant, which thus also supports the identity theory.

Our work has important implications for government policy. The standard policy prescription from the standard model on tax compliance to reduce tax evasion is an increase in audit probability or penalty rate or both. However, there are problems associated with the standard policy prescription. Our framework suggests a promising avenue for increasing tax compliance: measures that improve taxpayers' identification with taxpaying. It is important to point out that the new avenue helps to overcome the difficulties associated with standard policy prescription. Therefore, they tend to complement each other. Thus, an optimal and practical strategy might be to fully take advantage of both channels: improving taxpayers' identification with taxpaying while setting the audit probability and penalty rate at a reasonable level.

The remainder of the paper consists of five sections. The first section presents the theoretical model in detail by introducing the concept of identity into the standard model on tax compliance and shows that given an exogenous actual income, the more identified with taxpaying a taxpayer is, the more income he (or she) will declare to the tax authorities (i.e., there will be less tax evasion). The second and third sections present the empirical evidence from the 1987 Taxpayer Opinion Survey and the cross-state analysis respectively. Implications for government policy are discussed in the fourth section. The fifth section concludes.

THEORETICAL MODEL

In this section, we build a simple model by introducing the concept of identity into the standard model on tax compliance (Allingham and Sandmo, 1972) and show that given an exogenous actual income, the more identified with taxpaying a taxpayer is, the more income he (or she) will declare to the tax authorities (i.e., there will be less tax evasion).

The standard model on tax compliance is based on traditional expected utility theory. In this model, a rational individual takes his income (W) that is unknown to the tax authorities, the tax rate (θ), the audit probability (p), and the penalty rate (π) as given and chooses his declared income (X). After the individual declares his income, and if his declared income is less than his true income, he faces two possibilities: (1) With probability $(1-p)$, he will not be audited by the tax authorities so that he gains by $\theta(W-X)$; and (2) With probability p, he will be audited and the tax authorities will then know his true income. The consequence is that he will have to pay tax on the undeclared income $(W-X)$ at penalty rate (π) that is greater than tax rate (θ). In other words, he will lose by $(\pi-\theta)*(W-X)$.

The individual chooses his optimal declared income (X^*) by maximizing his expected utility function. Allingham and Sandmo assumed that the taxpayer's utility function has income as its only argument. Thus, the taxpayer will now choose X so as to maximize

$$\textbf{(1)} \quad E[u] = (1-p)u(W-\theta X) + pu(W-\theta X-\pi(W-X)).$$

Marginal utility, $u'(w)$, is assumed to be everywhere positive and strictly decreasing, so that the individual is risk averse. In addition, it is generally

assumed that the absolute risk aversion is decreasing with income. In other words, if we define $R_A(w) = -u''(w)/u'(w)$, then $R_A(w)$ is decreasing in w. Arrow (1970) gave a nice discussion of this measure.

Under some conditions satisfied by parameter values, Allingham and Sandmo showed that an interior solution would be guaranteed (i.e., $0 < X < W$). It is with such solutions that we shall be concerned in later analysis. The model implies that increasing audit probability (p) or penalty rate (π) can reduce tax evasion.

However, with identity taken into account, we instead assume that the taxpayer will now choose X so as to maximize

$$\textbf{(2)} \quad E[u] = (1-p)u(W - \theta X - I_c(W^c - X)) + pu(W - \theta X - \pi(W - X) - I_c(W^c - X)),$$

where c is the individual's social category, I_c is the individual's identity utility from being in category c and is positive, W^c is the ideal declared income in category c, and $I_c(W^c - X)$ is the disutility from diverging from the ideal declared income in category c -- the more identified with taxpaying, the higher I_c and W^c. Please note that in our model, we take I_c and W^c (i.e., people's identity), as given, which seems reasonable in the short run. However, when we discuss implications for government policy later, the objective of government policy is to improve people's identification with taxpaying. Similarly, it is also easy to show that under some conditions an interior solution will be guaranteed. In other words, we assume $0 < X < \min\{W^c, W\}$ following Allingham and Sandmo. We assume $\theta > I_c$ in our model because otherwise it is always better for the individual to declare $\min\{W^c, W\}$.

Equation (2) captures two important ideas not represented by utility functions of standard economics. First, it captures psychologists' and sociologists' view that decisions depend on social category. In the formula, the individual's utility varies with his category c. Second, the formula assumes that the individual's utility depends on the deviation of the declared income from the ideal declared income that varies with his category. These two points are not surprising because in a model of utility, a person's identity describes "gains and losses in utility from behavior that conforms or departs from the norms for particular social categories in particular situations." (Akerlof and Kranton, 2005). It is clear that our simple model is a variation of the standard model on tax compliance. Therefore, our simple model also illustrates that identity theory can be perfectly compatible with standard economic theory.

For notational convenience we define

$$\textbf{(3)} \quad Y = W - \theta X - I_c(W^c - X),$$

$$\textbf{(4)} \quad Z = W - \theta X - \pi(W - X) - I_c(W^c - X).$$

Clearly, by equations (3) and (4), we have $Y > Z$. As the general presumption is that the absolute risk aversion is decreasing with income, we have $R_A(Y) < R_A(Z)$.

The first-order condition for an interior maximum of equation (2) can then be written as

$$\textbf{(5)} \quad -(\theta - I_c)(1-p)u'(Y) - (\theta - \pi - I_c)pu'(Z) = 0.$$

The second-order condition

$$\textbf{(6)} \quad D = (\theta - I_c)^2(1-p)u''(Y) + (\theta - \pi - I_c)^2 pu''(Z) < 0$$

is satisfied by the assumption of concavity of the utility function.

Proposition 1 $\partial X^*/\partial I_c > 0$, where X^* is the optimal declared income that satisfies equation (5), the first-order condition.

Proof Take derivative with respect to I_c on both sides of equation (5) and by simple algebra manipulation, we have

$$\textbf{(7)} \quad \partial X^*/\partial I_c = (1-p)u'(Y) + pu'(Z) + (W^c - X^*)((\theta - I_c)(1-p)\,u''(Y) + (\theta - \pi - I_c)pu''(Z))/-D.$$

Substituting from equation (5) we can rewrite equation (7) as

$$\textbf{(8)} \quad \partial X^*/\partial I_c = (1-p)u'(Y) + pu'(Z) + (W^c - X^*)(\theta - I_c)(1-p) u'(Y)(R_A(Z) - R_A(Y))/-D.$$

As $-D > 0$, $(1-p)u'(Y) > 0$, $pu'(Z) > 0$, and $(W^c - X^*)(\theta - I_c)(1-p)u'(Y)(R_A(Z) - R_A(Y)) > 0$ implied by $(W^c - X^*) > 0$, $(\theta - I_c) > 0$, $(1-p)u'(Y) > 0$, and $(R_A(Z) - R_A(Y)) > 0$ because we assume absolute risk aversion is decreasing with income, the right-hand side of equation (8) is positive. Thus, we have proved that $\partial X^*/\partial I_c > 0$.

PROPOSITION 2 $\partial X^*/\partial W^c > 0$, where X^* is the optimal declared income that satisfies equation (5), the first-order condition.

Proof Take derivative with respect to W^c on both sides of equation (5) and by simple algebra manipulation, we have

$$\textbf{(9)} \quad \partial X^*/\partial W^c = I_c((\theta - I_c)(1-p)u''(Y) + (\theta - \pi - I_c)pu''(Z))/-D.$$

Substituting from equation (5) we can rewrite equation (9) as

$$\textbf{(10)} \quad \partial X^*/\partial W^c = I_c(\theta - I_c)(1-p)u'(Y)(R_A(Z) - R_A(Y))/-D.$$

As $-D > 0$, and $I_c(\theta - I_c)(1-p)u'(Y)(R_A(Z) - R_A(Y)) > 0$ implied by $I_c > 0$, $(\theta - I_c) > 0$, $(1-p)u'(Y) > 0$, and $(R_A(Z) - R_A(Y)) > 0$ because we assume absolute risk aversion is decreasing with income, the right-hand side of equation (10) is positive. Thus, we have proved that $\partial X^*/\partial W^c > 0$.

Interpretation: Taken together, as the more identified with taxpaying a taxpayer is, the higher I_c and W^c, proposition 1 and proposition 2 imply that given an exogenous actual income, the more identified with taxpaying a taxpayer is, the more income he (or she) will declare to the tax authorities (i.e., there will be less tax evasion).

EMPIRICAL EVIDENCE FROM THE 1987 TAXPAYER OPINION SURVEY

In this section, we analyze the 1987 Taxpayer Opinion Survey (TOS) data to empirically show that the more identified with taxpaying a taxpayer is, the less tax evasion there will be. Several other interesting findings will also be presented.

Data

The 1987 TOS Data utilized in this study were made available by the Inter-university Consortium for Political and Social Research. The data were originally collected by the Internal Revenue Service. There are a large number of questions in the survey and surprisingly not many researchers have used the data. Therefore, the data is attractive to us, and maybe to many other researchers.

The 1987 TOS provides taxpayers' opinions and evaluations of the U.S. tax system. Respondents were questioned about their knowledge of and feelings toward several recent tax reforms. They were also asked about their impression of the Internal Revenue Service and its programs, their experiences dealing with Internal Revenue Service agents, their opinions of the Internal Revenue Service's sharing information with other government agencies, and the sources of their information on taxes. In addition, attitudes towards tax evasion and towards those who cheat on their taxes were probed. Demographic information on each respondent was also collected.

The universe of the survey is adult taxpayers in the United States. The sampling method is stratified random sample. As taxpayers had sometimes the possibility to answer "Not Sure", or "Refuse" to answer some questions, the sample will be reduced depending on the number of control variables included in the regression if we set the value to be "missing" (i.e., discard the observation). As a robustness check, we also try an alternative method: setting the value to be the "median" of all possible effective answers.

Proxy for Identity (Independent Variable)

The answers to the following two questions are used as proxy for identity.

1. We'd like to talk to you about the income tax system. How do you feel about the federal income tax system as it applies to the 1986 tax return – do you feel it is quite fair to most people, or reasonably fair, or somewhat unfair, or quite unfair to most people?

   Quite Fair ……..….—— 4
   Reasonably Fair …….—— 3
   Somewhat Unfair……—— 2
   Quite Unfair ……...—— 1
   Don't Know ……...—— 5

2. Public Officials can usually be trusted to do what's right

   Strongly Agree……—— 4
   Mildly Agree……...—— 3
   Mildly Disagree……..—— 2
   Strongly Disagree …..—— 1
   Not Sure …… —— 5

Why do we use the answers to the above two questions as proxy for identity? The argument is obvious: The more fair the taxpayer thinks the tax system is, the more identified with taxpaying he will be; The more the taxpayer trusts in the public (especially tax) officials, the more identified with taxpaying he will be.

For notional convenience, we define the answers to the above two questions as *Fairness_Identity* and *Trust_Identity*, respectively. We further define *Identity* = (*Fairness_Identity* + *Trust_Identity*)/2 as proxy for identity.

To be consistent with our theoretical model, we recode (Quite Fair, Reasonably Fair, Somewhat Unfair, Quite Unfair) from (1, 2, 3, 4) to (4, 3, 2, 1). Similarly, we recode (Strongly Agree, Mildly Agree, Mildly Disagree, Strongly Disagree) from (1, 2, 3, 4) to (4, 3, 2, 1). If taxpayer answers "Not Sure" or "Don't Know", we use two methods to deal with it: 1) setting the value to be "missing", i.e., we discard this observation; and 2) setting the value to be 2.5, the median of all possible effective answers: (1, 2, 3, 4).

Proxy for Tax Evasion (Dependent Variable)

The answers to the following four questions are used as proxy for tax evasion.

1. Trading or exchanging goods or serv-ices with a friend or neighbor and not reporting it on your tax form.

   (Perfectly Acceptable = 6 5 4 3 2 1 = Not at all Acceptable; 9=Not Sure)

2. Reporting your main income fully, but not including some small outside income.

   (Perfectly Acceptable = 6 5 4 3 2 1 = Not at all Acceptable; 9=Not Sure)

3. Being paid in cash for a job and then not reporting it on your tax form.

   (Perfectly Acceptable = 6 5 4 3 2 1 = Not at all Acceptable; 9=Not Sure)

4. Not reporting some earnings from invest-ments or interest that the government would not be able to find about.

   (Perfectly Acceptable = 6 5 4 3 2 1 = Not at all Acceptable; 9=Not Sure)

Why do we use the answers to the above four questions as proxy for tax evasion? The argument is also obvious: The more acceptable the taxpayer thinks the above practices are, the more likely he will do it and thus the more tax evasion.

Let *TaxEvasion*_1, *TaxEvasion*_2, *TaxEvasion*_3 and *TaxEvasion*_4 be the answers to the above four questions. We further define *TaxEvasion* = $\Sigma_{i=1}^{4}$*TaxEvasion_i*/4 as proxy for tax evasion.

Similarly, if taxpayer answers "Not Sure", we use two methods to deal with it: (1) setting the corresponding value to be "missing" (i.e., discard the observation); and (2) setting the value to be 3.5, the median of all possible effective answers: (1, 2, 3, 4, 5, 6).

Control Variables

We include seven demographic variables as control variables: a dummy variable for female (coded 1 for females and 0 for males), age, a dummy variable for white people (coded 1 for white people and 0 for other races), a dummy variable for single marital status (coded 1 for single and 0 for others such as Married, Separated, Divorced, etc.), a dummy variable for employed people (coded 1 for working full-time, working part-time, working without pay on a family farm or business and with a job but not at work because of temporary illness, vacation, strike, etc., and 0 for others), education, and income.

Econometric Results

As we mentioned before, if taxpayer answers "Not Sure" or "Don't Know" or refuses to answer the question, we use two methods to deal with it: (1) setting the value to be "missing" (i.e., we discard this observation); and (2) setting the value to be the median of all possible effective answers. Table 1 reports the results using the first method, and Table 2 reports the results using the second method.

First, let's take a look at Table 1. As a rough check, we first regress *TaxEvasion* on *Identity*. Column 1 of Table 1 presents the estimation results. From column 1 of Table 1, we can see that the coefficient of *Identity* is –0.12621 and statistically significant because its p-value is only 0.0058. The negative coefficient of *Identity* means that the more identified with taxpaying a taxpayer is, the less tax evasion. Therefore, the rough check supports our theoretical conclusion.

Table 1
Report for Linear Regression Results
(Data Set: The 1987 Taxpayer Opinion Survey)

	Dependent Variable: TaxEvasion			
Independent Variables	*Column 1*	*Column 2*	*Column 3*	*Column 4*
Intercept	3.27536 (SE: 0.10824) (P: <.0001)	4.64030 (SE: 0.25402) (P: <.0001)	4.79351 (SE: 0.15912) (P: <.0001)	5.03507 (SE: 0.17218) (P: <.0001)
Identity	-0.12621 (SE: 0.04566) (P: 0.0058)	-0.17374 (SE: 0.04704) (P: 0.0002)	-0.14713 (SE: 0.04384) (P: 0.0008)	
Fairness_Identity				-0.09555 (SE: 0.03982) (P: 0.0165)
Trust_Identity				-0.14841 (SE: 0.03834) (P: 0.0001)
Female		-0.26489 (SE: 0.07389) (P: 0.0003)	-0.28403 (SE: 0.06685) (P: <.0001)	-0.24904 (SE: 0.06970) (P: 0.0004)
Age		-0.02328 (SE: 0.00274) (P: <.0001)	-0.02358 (SE: 0.00196) (P: <.0001)	-0.02252 (SE: 0.00208) (P: <.0001)
White		-0.24975 (SE: 0.10372) (P: 0.0162)	-0.31259 (SE: 0.09432) (P: 0.0009)	-0.34855 (SE: 0.09874) (P: 0.0004)
Single		-0.02727 (SE: 0.10829) (P: 0.8012)		
Employed		-0.03171 (SE: 0.10038) (P: 0.7521)		
Education		0.02235 (SE: 0.02313) (P: 0.3342)		
Income		0.00555 (SE: 0.01803) (P: 0.7581)		

Note: If taxpayer answers "Not Sure" or "Don't Know" or refuses to answer the question, we set the value to be "missing."

We then include all the seven control variables and regress *TaxEvasion* on *Identity*, female, age, white, single, employed, education, and income. Column 2 of Table 1 presents the estimation results. From column 2 of Table 1, we can see that the coefficient of *Identity* is –0.17374 and statistically significant because its p-value is only 0.0002. The negative coefficient of *Identity* means that the more identified with taxpaying a taxpayer is, the less tax evasion. Therefore, the empirical evidence supports our theoretical conclusion.

Table 2
Report for Linear Regression Results
(Data Set: The 1987 Taxpayer Opinion Survey)

	Dependent Variable: TaxEvasion			
Independent Variables	*Column 1*	*Column 2*	*Column 3*	*Column 4*
Intercept	3.71880 (SE: 0.11848) (P: <.0001)	5.04368 (SE: 0.25171) (P: <.0001)	5.03176 (SE: 0.15893) (P: <.0001)	5.02957 (SE: 0.15929) (P: <.0001)
Identity	-0.26050 (SE: 0.04814) (P: <.0001)	-0.27836 (SE: 0.05002) (P: <.0001)	-0.24734 (SE: 0.04639) (P: <.0001)	
Fairness_Identity				-0.11711 (SE: 0.03777) (P: 0.0020)
Trust_Identity				-0.12963 (SE: 0.03564) (P: 0.0003)
Female		-0.2346 (SE: 0.07020) (P: 0.0009)	-0.22567 (SE: 0.06359) (P: 0.0004)	-0.22525 (SE: 0.06384) (P: 0.0004)
Age		-0.02086 (SE: 0.00260) (P: <.0001)	-0.02113 (SE: 0.00186) (P: <.0001)	-0.02111 (SE: 0.00187) (P: <.0001)
White		-0.24303 (SE: 0.09807) (P: 0.0133)	-0.32946 (SE: 0.08921) (P: 0.0002)	-0.32997 (SE: 0.08927) (P: 0.0002)
Single		-0.00034 (SE: 0.10277) (P: 0.9973)		
Employed		-0.03155 (SE: 0.09475) (P: 0.7392)		
Education		0.00488 (SE: 0.02204) (P: 0.8249)		
Income		-0.00624 (SE: 0.01719) (P: 0.7168)		

Note: If taxpayer answers "Not Sure" or "Don't Know" or refuses to answer the question, we set the value to be the median of all possible effective answers.

Among the seven control variables, female, age, and white are statistically significant because their p-values are 0.0003, less than 0.0001, and 0.0162, respectively; while single, employed, education, and income are not because their p-values are 0.8012, 0.7521, 0.3342, and 0.7581, respectively. Therefore, we rerun the regression including only the significant control variables. In other words, we regress *TaxEvasion* on *Identity*, female, age, and white. Column 3 of Table 1 presents the estimation results. From column 3 of Table 1, we can see that the coefficient of *Identity* is –0.14713 and

statistically significant because its p-value is only 0.0008. The negative coefficient of *Identity* means that the more identified with taxpaying a taxpayer is, the less tax evasion. Therefore, the empirical evidence supports our theoretical conclusion. All the three control variables are still negative (-0.28403, -0.02358, and –0.31259, respectively) and statistically significant because their p-values are less than 0.0001, less than 0.0001, and 0.0009, respectively. Therefore, females tend to comply with the law, while males not; older people tend to comply with the law, while younger people not; white people tend to comply with the law, while other races not.

We define *Identity* = (*Fairness_Identity* + *Trust_Identity*)/2 as proxy for identity. To see the correlation between *Fairness_Identity* and *TaxEvasion*, and the correlation between *Trust_Identity* and, *TaxEvasion* respectively; we regress *TaxEvasion* on *Fairness_Identity*, *Trust_Identity*, female, age, and white. Column 4 of Table 1 presents the estimation results. From Column 4 of Table 1, we can see that the coefficients of *Fairness_Identity* and *Trust_Identity* are –0.09555 and –0.14841, respectively and are both statistically significant because their p-values are 0.0165 and 0.0001, respectively. Therefore, both *Fairness_Identity* and *Trust_Identity* are negatively correlated with *TaxEvasion*, but the coefficient of *Trust_Identity* is more negative. Similarly, the coefficients of female, age, and white are still negative (-0.24904, -0.02252, and –0.34855, respectively) and statistically significant because their p-values are 0.0004, less than 0.0001, and 0.0004, respectively. Therefore, females tend to comply with the law, while males not; older people tend to comply with the law, while younger people not; white people tend to comply with the law, while other races not.

Now, let's take a look at Table 2 that reports the results using the second method to deal with "Not Sure" or "Don't Know." Comparing Table 2 with Table 1, we can see that the basic conclusions do not change:

1. The coefficient of *Identity* is both negative and statistically significant, which means that the more identified with taxpaying a taxpayer is, the less tax evasion there will be;
2. Gender, age, and race are significant factors affecting tax evasion: females tend to comply with the law, while males not; older people tend to comply with the law, while younger people not; white people tend to comply with the law, while other races not;
3. Both *Fairness_Identity* and *Trust_Identity* are negatively correlated with *TaxEvasion*, but estimated coefficient of *Trust_Identity* is more negative.

Brief Summary and Discussions

The results of the eight regressions we run so far are somewhat consistent. First, these results support our theoretical conclusion: The more identified with taxpaying a taxpayer is, the less tax evasion there will be.

Second, we find that females, older people, and white people tend to comply with the law, while others do not. These three findings are consistent with Roth et al. (1989). However, our analysis cannot tell us whether taxpayers grow more compliant as they age ("aging effect"), or whether successive cohorts of taxpayers are becoming less compliant ("cohort effect"). Distinguishing between aging effect and cohort effect is essential for predicting future compliance levels. If data permitted, it should be very interesting to pursue in this direction.

EMPIRICAL EVIDENCE FROM CROSS-STATE ANALYSIS

One criticism of the approach employed to analyze the 1987 TOS data is that it might really study the relationship between views on tax compliance and views of goods, not actual tax compliance. This concern drives us to go beyond survey data analysis and do a cross-state analysis. In this section, we present our cross-state analysis including the econometric model, the data, and the result.

Econometric Model

Before presenting our econometric model, it is worth reviewing the literature on tax compliance taking a cross-section approach at aggregate level. Perhaps due to limited data availability, only a few papers study tax compliance at aggregate level. Witte and Woodbury (1985), Dubin and Wilde (1988), and Beron et al. (1992) all investigated the TCMP data for tax year 1969 aggregated by the IRS to the 3-digit zip code level. This data set includes an estimated tax compliance variable, agency variables such as audit rates, and demographic variables aggregated to the 3-digit zip code level.

The methodology they employed in their studies is linear regression analysis. Witte and Woodbury (1985) first analyzed this data set. They found that there is a significant negative relationship between audit rate and tax evasion. However, they failed to recognize the potential endogeneity of audit rate, which is implied by the game theoretic literature (Reinganum and Wilde, 1985; Border and Sobel, 1987). The game-theoretic literature assumes that the IRS acts strategically and can adjust its audit rules based on its information on taxpayers. Thus, the game-theoretic literature implies that audit rate is endogenous. Dubin and Wilde (1988) recognized this problem and investigated this data set employing the instrumental variable approach. They used the IRS budget per tax return filed as the instrumental variable for audit rate. They argued that this is a valid instrumental variable because they analyzed the time path of state-level IRS budgets and found them to be independent of compliance levels. Instead, they found that the state-level IRS budgets are predominately determined by the share of total return filed. However, their conclusion is consistent with that of Witte and Woodbury (1985): Even after allowing for the endogeneity of audit rate, they still found that auditing has a significant deterrent effect on tax evasion. Beron et al. (1992) reanalyzed the same data set. This time they used the total number of returns filed in a district divided by the number of full-time equivalent IRS district employees as the instrumental variable for audit rate. Essentially, this instrument is the inverse of the instrument employed by Dubin and Wilde (1985). They also found that auditing has a positive effect on tax compliance. Both Dubin and Wilde (1985) and Beron et al. (1992) estimate their model using the methodology of 2-stage least squares (2SLS). At state level, we only find three papers in the literature: Dubin et al. (1987, 1990) and Plumley (1996). Dubin et al. (1987, 1990) combined data from the IRS Annual Report of the Commissioner with the Statistics of Income to create a panel data set for the states from 1977 through 1986. They used IRS budget per return filed and the number of information returns other than W2 forms per return filed as instruments for audit. Their conclusion is that the continual decline of audit rate has a significant negative effect on IRS tax collections. However, the drawback of their work is that they don't have a direct measure of tax compliance. Instead, they used returns filed per capita (total individual income tax returns filed divided by total population) and reported taxes per return (reported individual income tax divided by the number of individual tax returns filed) as their measures of tax compliance. Plumley (1996) followed a similar approach as Dubin et al. (1987, 1990). He created a panel data set for the states from 1982 through 1991 based on numerous IRS and Census data sources. His model consists of four compliance equations and one first-stage audit rate equation. He found that many tax policy instruments such as audits have significant effects on tax compliance.

Now, we come back to our own cross-state analysis. Like the literature on tax compliance taking a cross-section approach at aggregate level (either at 3-digit zip code level or at state level), we employ the methodology of linear regression analysis. Because we are only interested in the impact of identity on tax evasion and in order to avoid the trouble of possible endogeneity of state-level audit rate, we employ the following econometric model in our study:

$$\textbf{(11)} \quad StateLevelTaxEvasion = \alpha + \beta * StateLevelIdentity + \varepsilon,$$

where ε is a random variable, *StateLevelTaxEvasion* measures state-level tax evasion, and *StateLevelIdentity* measures statewide identification with taxpaying. In the long run, state-level identity should be an endogenous variable. But in the short run, it is reasonable to argue that state-level identity is exogenous and stable. In this research, we take state-level identity as exogenous and stable since we are doing a cross-state analysis. If the estimated coefficient of *StateLevelIdentity*, i.e., β, is negative and statistically significant, then the result of the cross-state analysis is consistent with the identity theory.

The advantage of this econometric model is that it focuses exclusively on identity and leaves all other possible factors to the residual and thus avoids the trouble of a possible endogenous state-level audit rate.

Measure of State-Level Tax Evasion

StateLevelTaxEvasion is a measure of state-level tax evasion. In this research, we use the state-level Wage and Salary Gap (the proportion of wage and salary income underreported) for year 2000 produced by the U.S. Bureau of Economic Analysis (BEA) as *StateLevelTaxEvasion* (Brown et al., 2006). The decision to limit the analysis to

wage and salary income underreporting is made for several reasons. First, wage and salary income is the most frequently reported category of income on tax returns. Second, this tax evasion measure is relatively reliable. It is worth mentioning that the total adjusted gross income (AGI) gap also produced by BEA has been used as a compliance measure by several researchers (Crane and Nourzad, 1986; Engel and Hines, 1998), but its reliability as a measure of tax evasion is questionable. This is because a significant share of the total AGI gap is due to nonfarm proprietors' income that relies on data from individual tax returns. In contrast, the wage and salary component of the AGI gap uses data by employers to the U.S. Bureau of Labor Statistics and thus is more reliable. However, it is important to point out that the BEA Wage and Salary Gap measure includes income not subject to tax and therefore should be considered only as a "rough" measure of tax evasion. Finally, this is a relatively new data set. We are not aware that this data set has been used by other researchers.

Measure of State-Level Identity

StateLevelIdentity measures statewide identification with taxpaying. In order to do the cross-state analysis, we need to construct an index for state-level identity. The criterion of a good index for state-level identity is that the index should capture statewide attitude toward taxpaying. Guided by this criterion, we have investigated several possible data sources to construct the index for state-level identity.

First, we have investigated the 1987 TOS. In the 1987 TOS, survey respondents are asked about their attitude toward tax evasion, tax systems, and the government (especially IRS). For a detailed description of the 1987 TOS, please see the previous section. From the 1987 TOS data, we construct the proxy for individual identity (from Question 1a and Question 103b) in the previous section. Because we can also extract the state information for each survey respondent, a natural way to construct the index for state-level identity is to take the average of individual identity for each state as *StateLevelIdentity*. However, there are two concerns. First, the sample size for some states is very small (as less as 10) so that such an index might contain much noise. This concern can be addressed by our empirical strategy, which will be described later. Second, *StateLevelTaxEvasion* and *StateLevelIdentity* are constructed from year 2000 data and year 1987 data, respectively so that *StateLevelIdentity* from year 1987 might not be still applicable in year 2000. There appears to have been erosion in public attitudes toward taxpaying. But as we are doing cross-state analysis, this concern may not have much impact. We denote the state-level identity constructed from the 1987 TOS as *StateLevelIdentity_TOS*1987.

Second, we have investigated The Rush Limbaugh Show. The Rush Limbaugh Show is a right-wing radio show and is anti-tax. The popularity of this show in each state may capture a statewide attitude that is correlated with statewide attitude toward taxpaying. Therefore, the popularity of this show could be a good index for state-level identity. However, such data is unavailable due to the way in which the radio network market works.

Third, we have investigated the voting and registration rates in the November 2000 U.S. presidential election. The right to vote is arguably one of the most important rights of citizenship in a democratic country. We argue that an individual's deciding to register and vote indicates that he (or she) is civic-minded. Thus, registration and voting might be correlated with his (or her) attitude toward taxpaying. Following this argument, the higher the voting and registration rates for a state, the higher the state-level identity. From the 2000 Current Population Survey (CPS), the U.S. Census Bureau has constructed for each state the registration and voting rates based on both the voting-age population and the citizen population of voting age. Correspondingly, we construct four indexes for state-level identity:

1. $StateLevelIdentity\_1 = \frac{TotalRegistered}{Population \geq 18}$;
2. $StateLevelIdentity\_2 = \frac{TotalVoted}{Population \geq 18}$;
3. $StateLevelIdentity\_3 = \frac{TotalRegistered}{TotalCitizen \geq 18}$;

   and
4. $StateLevelIdentity\_4 = \frac{TotalVoted}{TotalCitizen \geq 18}$.

Such data is tabled and available at the Web site of the U.S. Census Bureau (http://www.census.gov/population/www/socdemo/voting.html). In addition, the four indexes for state-level identity and the measure for state-level tax evasion are all from year 2000.

Finally, we argue that there might be a strong correlation between statewide identification with taxpaying and state-level murder rate because, in a more general sense, both are correlated with statewide identification with legal systems. Following this argument, the higher the state-level murder rate, the lower the statewide identification with taxpaying. Therefore, we construct one more index for state-level identity: *StateLevelIdentity_MurderRate* = 1/*StateLevelMurderRate*_2000 (i.e., the index for state-level identity is the inverse of state-level murder rate in year 2000). Thus, if we regress *StateLevelTaxEvasion* on *StateLevel Murderrate*_2000, we expect that the estimated coefficient of *StateLevelMurderrate*_2000 is positive. *StateLevelMurderrate*_2000 is tabled and available at the Death Penalty Information Center (http://www.deathpenaltyinfo.org).

In summary, we have constructed six indexes for state-level identity: *StateLevelIdentity_TOS*1987 from the 1987 TOS, *StateLevelIdentity*_1, *StateLevelIdentity*_2, *StateLevelIdentity*_3, and *StateLevelIdentity*_4 from the voting and registration rates in the November 2000 U.S. presidential election, and finally *StateLevelIdentity_MurderRate* from the state-level murder rate in year 2000.

Results of Cross-State Analysis

First, we regress *StateLevelTaxEvasion* on *StateLevelIdentity_TOS*1987. We find that the estimated coefficient of *StateLevelIdentity_TOS*1987 is negative (-0.02025) and statistically significant because the p-value is around 0.07. The negative coefficient of *StateLevelIdentity_TOS*1987 means that the more identified with taxpaying, the less tax evasion there will be. Thus, this result is consistent with the identity theory. Just as mentioned before, one concern is that the sample size for some states is very small (as less as 10) so that *StateLevelIdentity_TOS*1987 might contain much noise. In order to address this concern, we take the following strategy. First, for each state, we randomly split the correspondents into two even groups. For the two groups, we construct *StateLevelIdentity_TOS*1987_*Group*1 and *StateLevelIdentity_TOS*1987_*Group*2, respectively, by taking the same approach to construct *StateLevelIdentity_TOS*1987. Then we use *StateLevelIdentity_TOS*1987_*Group*2 as the Instrumental Variable (IV) for *StateLevelIdentity_TOS*1987_*Group*1 and do cross-state regression. We find that the estimated coefficient of *StateLevelIdentity_TOS*1987_*Group*1 is negative (-0.027434) and statistically significant because the p-value is 0.023. Symmetrically, we also use *StateLevelIdentity_TOS*1987_*Group*1 as the instrumental variable (IV) for *StateLevel Identity_TOS*1987_*Group*2 and do cross-state regression. We find that the estimated coefficient of *StateLevelIdentity_TOS*1987_*Group*2 is negative (-0.0271495) and statistically significant because the p-value is 0.019. Thus, the results of the two IV regressions also support the identity theory. The results of these three regressions are reported in columns 1, 2, and 3 of Table 3, respectively.

Second, we regress *StateLevelTaxEvasion* on *StateLevelIdentity*_1, *StateLevelIdentity*_2, *StateLevelIdentity*_3, and *StateLevelIdentity*_4, respectively. The results of these four regressions are reported in columns 1, 2, 3, and 4 of Table 4, respectively. Columns 1, 2, 3, and 4 of Table 4 show that the estimated coefficients of *StateLevel Identity*_1, *StateLevelIdentity*_2, *StateLevel Identity*_3, and *StateLevelIdentity*_4 are all negative (-0.09794, -0.14812, -0.08957, and –0.14382, respectively) and statistically significant because their p-values are 0.0186, 0.0017, 0.0721, and 0.0084, respectively. The negative coefficients of *StateLevelIdentity*_1, *StateLevelIdentity*_2, *StateLevelIdentity*_3, and *StateLevelIdentity*_4 mean that the more identified with taxpaying, the less tax evasion there will be. Therefore, the results of these four regressions also support the identity theory.

Finally, we regress *StateLevelTaxEvasion* on *StateLevelMurderrate*_2000, the inverse of *StateLevelIdentity_MurderRate*. Just as expected, the estimated coefficient of *StateLevel Murderrate*_2000 is positive (0.00335) and statistically significant because the p-value is only 0.0065. Thus, the result of this regression supports the identity theory.

In summary, the results of all the eight cross-state regressions significantly support the identity theory: the more identified with taxpaying, the less tax evasion there will be.

IMPLICATIONS FOR GOVERNMENT POLICY

So far, we have shown, both theoretically and empirically, that identity plays an important role in taxpayers' tax compliance decision. We believe our work has important implications for government policy.

Table 3
Report for Cross-State Regression
(State-level Identity Constructed from the 1987 Taxpayer Opinion Survey)

Dependent Variable: StateLevelTaxEvasion			
Independent Variables	*Column 1*	*Column 2*	*Column 3*
Intercept	0.10839 (SE: 0.02553) (P: 0.0001)	0.12492 (SE: 0.02692) (P: <.0001)	0.12405 (SE: 0.02527) (P: <.0001)
*StateLevelIdentity_TOS*1987	-0.02025 (SE: 0.01111) (P: 0.0756)		
*StateLevelIdentity_TOS*1987_*Group*1 (*IV :StateLevelIdentity_TOS*1987_*Group2*)		-0.027434 (SE: 0.01164) (P: 0.023)	
*StateLevelIdentity_TOS*1987_*Group2* (*IV :StateLevelIdentity_TOS*1987_*Group1*)			-0.0271495 (SE: 0.01115) (P: 0.019)

Table 4
Report for Cross-State Regression
(State-level Identity Constructed from the 2000 U.S. Presidential Election)

Dependent Variable: StateLevelTaxEvasion				
Independent Variables	*Column 1*	*Column 2*	*Column 3*	*Column 4*
Intercept	0.12575 (SE: 0.02705) (P: <.0001)	0.14534 (SE: 0.02570) (P: <.0001)	0.12320 (SE: 0.03436) (P: 0.0008)	0.14709 (SE: 0.03172) (P: <.0001)
StateLevelIdentity_1	-0.09794 (SE: 0.04021) (P: 0.0186)			
StateLevelIdentity_2		-0.14812 (SE: 0.04447) (P: 0.0017		
StateLevelIdentity_3			-0.08957 (SE: 0.04870) (P: 0.0721)	
*StateLevelIdentity*_4				-0.14382 (SE: 0.05232) (P: 0.0084)

Governments everywhere wish to increase tax compliance. By the standard model on tax compliance, the standard policy prescription to reduce tax evasion is an increase in audit probability or penalty rate or both. Much empirical evidence has confirmed that greater enforcement can increase compliance (Dubin et al., 1990). However, there are problems associated with the standard policy prescription and it is unlikely that the extreme measures will be implemented (Skinner and Slemrod, 1985). First, serious penalties would cause a large divergence in treatment between successful evaders

and those who were gathered up in the IRS random audits. Higher penalties would thus violate ex post horizontal equity by punishing a few taxpayers as examples to others. Second, taxpayers do make honest mistakes, failing to declare income that they believe is deductible, or neglecting to keep records for legitimate expenses. In the vast majority of tax evasion cases, it is difficult for the tax authority to distinguish honest from dishonest mistakes. Given the complexity of the tax code, and the multitude of chances for error, severe punishments for "good faith" filing errors seem inequitable. Finally, there is a widespread belief that "the punishment should fit the crime."

Compared with the standard policy prescription, our framework suggests a promising avenue for reducing tax evasion: measures that improve taxpayers' identification with taxpaying. Clearly, our policy suggestion helps to overcome the difficulties associated with the standard practices: For a given level of tax compliance, if taxpayers are more identified with taxpaying, the audit probability and penalty rate could be reduced so that the problems associated with the standard practices can at least be alleviated.

The above discussion implies that our approach and the standard practices tend to complement each other. Thus, an optimal and practical strategy might be to fully take advantage of both channels: increasing taxpayers' identification with taxpaying while setting the audit probability and penalty rate at a reasonable level.

CONCLUSION

In this research, we are relating tax evasion to taxpayers' identity. Theoretically, we build a simple model showing that the more identified with taxpaying a taxpayer is, the less tax evasion there will be. Empirically, both the analysis of the 1987 Taxpayer Opinion Survey data and the cross-state analysis are consistent with the identity theory.

Compared with the standard policy prescriptions to reduce tax evasion, such as increasing audit probability and penalty rates to reduce tax evasion, our work suggests a promising avenue for increasing tax compliance: measures that improve taxpayers' identification with taxpaying. Our policy suggestion overcomes the difficulties associated with the standard practices. In addition, our policy suggestion and standard practices tend to complement each other, which implies that an optimal and practical strategy might be to fully take advantage of both channels (i.e., increasing taxpayers' identification with taxpaying while setting the audit probability and penalty rate at a reasonable level).

References

Akerlof, George A. and Rachel E. Kranton.
Economics and Identity. *Quarterly Journal of Economics* (August 2000): 715-753.
Identity and Schooling: Some Lessons for the Economics of Education. *Journal of Economic Literature* 40 (December 2002): 1167-1201.
Identity and the Economics of Organizations. *Journal of Economic Perspectives* 19 (2005): 9-32.

Allingham, Michael G. and Agnar Sandmo. Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics* 1 (November 1972): 323-338.

Arrow, Kenneth J. *Essays in the Theory of Risk-Bearing*. Amsterdam: North-Holland Publishing Co., 1970.

Beron, Kurt J., Helen V. Tauchen, and Ann Dryden Witte. The Effect of Audits ad Socioeconomic Variables on Compliance. In Joel Slemrod, eds. *Why People Pay Taxes*. Ann Arbor: University of Michigan Press, 1992.

Border, Kim and Joel Sobel. Samurai Accountant: A Theory of Auditing and Plunder. *Review of Economic Studies* 54 (October 1987): 525-540.

Brown, Robert L., Ann E. Dunbar, and Adrienne T. Pilot. The Feasibility of Producing Personal Income to Adjusted Gross Income (PI-AGI) Reconciliations by State. Washington, D.C.: Bureau of Economic Analysis, 2006. BEA Working Paper WP2006-05.

Crane, Steven E. and Farrokh Nourzad. Inflation and Tax Evasion: An Empirical Analysis. *Review of Economics and Statistics* 68 (May 1986): 217-223.

Dubin, Jeffrey A., Michael J. Graetz, and Louis L. Wilde.
Are We a Nation of Tax Cheaters? New Econometric Evidence on Tax Compliance. *American Economic Review* 77 (May 1987): 240-245.
The Effects of Audit Rates on the Federal Individual Income Tax, 1977-1986. *National Tax Journal* 43 (December 1990): 395-409.

Dubin, Jeffrey A. and Louis L. Wilde. An Empirical Analysis of Federal Income Tax Auditing and Compliance. *National Tax Journal* 41 (March 1988): 61-74.

Engel, Eduardo M.R.A. and James R. Hines Jr. Understanding Tax Evasion Dynamics. Ann Arbor, MI: Office of Tax Policy Research, University of Michigan, 1998. OTPR Working Paper 98-22.

Harris, Louis and Associates, Inc. 1987 Taxpayer Opinion Survey. Washington, D.C.: Conducted for the U.S.

Department of Treasury, Internal Revenue Service, 1988. IRS Document 7292.

Plumley, Alan H. The Determinants of Individual Income Tax Compliance: Estimating the Impacts of Tax Policy, Enforcement, and IRS Responsiveness. Washington, D.C.: Internal Revenue Service, 1996. IRS Publication 1916.

Reinganum, Jennifer F. and Louis L. Wilde. Income Tax Compliance in a Principal-Agent Framework. *Journal of Public Economics* 26 (1985): 1-18.

Roth, Jeffrey A., John T. Scholz, and Ann D. Witte, eds. *Tax Compliance: An Agenda for Research*. Philadelphia: University of Pennsylvania Press, 1989.

Skinner, Jonathan and Joel Slemrod. An Economic Perspective on Tax Evasion. *National Tax Journal* 38(3) (1985): 345-353.

Witte, Ann D. and Diane F. Woodbury. The Effect of Tax Laws and Tax Administration on Tax Compliance: The Case of the U.S. Individual Income Tax. *National Tax Journal* 38 (March 1985): 1-13.