Dear Author,

During the preparation of your manuscript for typesetting some questions have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin of the proof or compile them as a separate list. This form should then be returned with your marked proof/list of corrections to Elsevier Science.

**Disk use**
In some instances we may be unable to process the electronic file of your article and/or artwork. In that case we have, for efficiency reasons, proceeded by using the hard copy of your manuscript. If this is the case the reasons are indicated below:

- Disk damaged
- Incompatible file format
- Virus infected
- Discrepancies between electronic file and (peer-reviewed, therefore definitive) hard copy.
- Other: .................................................................

We have proceeded as follows:
- Manuscript scanned
- Manuscript keyed in
- Artwork scanned
- Files only partly used (parts processed differently: .................................................................)

**Bibliography**
If discrepancies were noted between the literature list and the text refer, the following may apply:

- The references listed below were noted in the text but appear to be missing from your literature list. Please complete the list or remove the references from the text.

- **Uncited references:** This section comprises references which occur in the reference list but not in the body of the text. Please position each reference in the text or, alternatively, delete it. Any reference not dealt with will be retained in this section.

**Queries and/or remarks**

<table>
<thead>
<tr>
<th>Manuscript page/line</th>
<th>Details required</th>
<th>Author’s response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Please supply full addresses for affiliations 'b' and 'c'. The following references are quoted in the text but do not appear in the ref. list: Section 2, first sentence - Allingham and Sandmo, 1972 (also Sections 5.2.1 and 5.2.2); Yitzhaki, 1974 (also Section 5.2.2)</td>
</tr>
</tbody>
</table>
The determinants of capital gains tax compliance: evidence from the RJR Nabisco leveraged buyout

Wayne R. Landsman\(^8\), Douglas A. Shackelford\(^{a,b,*}\), Robert J. Yetman\(^c\)

\(^{a}\)Kenan-Flagler Business School, University of North Carolina, Campus Box 3490, McColl Building, Chapel Hill, NC 27599-3490, USA
\(^{b}\)NBER, city, USA
\(^{c}\)University of Iowa, city, USA

Received 5 June 2000; accepted 18 January 2001

---

Abstract

Inability to observe both actual and reported capital gains has impeded studies of capital gains tax compliance. This study overcomes such limitations through access to both confidential tax returns and internal shareholder records. Tests are conducted at the individual taxpayer level for associations between compliance rates for capital gains taxes generated by RJR Nabisco shareholders during its leveraged buyout and income, marginal tax rates, socioeconomic characteristics, and measures of the taxpayer's perception of the noncompliance penalty. We find noncompliance decreasing in taxable income, including RJR Nabisco capital gains, first-dollar marginal tax rates, self-employment income, rental income, and other interest paid. © 2001 Published by Elsevier Science B.V.

Keywords: Capital gains tax compliance; Evasion; Acquisitions

JEL classification: H24; H26; G34

---

1. Introduction and motivation

Despite the importance of capital gains taxes as a source of individual tax

---

\(^8\)Corresponding author. Kenan-Flagler Business School, University of North Carolina, Campus Box 3490, McColl Building, Chapel Hill, NC 27599-3490, USA. Tel.: +1-919-962-3197; fax: +1-919-962-4727.

E-mail address: doug.shack@unc.edu (D.A. Shackelford).
revenue and their documented understatement on tax returns, little is known about
the determinants of capital gains tax compliance. This omission in the literature
exists because researchers have been unable to observe both actual and reported
capital gains. This study overcomes this hurdle through access to both confidential
individual tax returns, provided by the Internal Revenue Service, and confidential
shareholder records, provided by RJR Nabisco. These data enable us to test for
associations among the socioeconomic characteristics of individual shareholders,
their tax return information, including marginal tax rates, and their compliance
rates for the capital gains taxes generated by the 1989 RJR Nabisco leveraged
buyout (LBO).

Capital gains taxes are assessed at realization on the difference between sales
proceeds and the investor’s tax basis in the disposed property. Although theory
(Constantinides, 1983; Stiglitz, 1983; Scholes and Wolfson, 1992, among others)
suggests that perfect substitutability among financial assets could preclude capital
gains taxes, individuals continue to report sizable capital gains. For example, in
1997, even after deducting capital losses, net capital gains comprised 7 percent of
adjusted gross income (AGI), trailing only wages as an income source (IRS
Statistics of Income Files). This percentage has doubled since 1991, and the
sustained bull market in the 1990s portends an increasing importance of capital
gains as a revenue source.

Moreover, the limited publicly available data suggest that taxpayers realize
more capital gains than they report on their tax returns. IRS estimates from the
nine Taxpayer Compliance Measurement Program (TCMP) surveys from 1965 to
1988 indicate that the percent of actual capital gains that appear on tax returns
ranges from 61 percent in 1976 to 93 percent in 1988.¹ This compares with 99
percent compliance for wages and 98 percent for interest in the 1988 TCMP
survey.

Several factors aid capital gains noncompliance. Capital gains transactions are
irregular and infrequent, making it difficult for the taxing authorities to establish
baselines cross-temporally or cross-taxpayers. In addition, tax bases are not
tracked by third-party reports, and unless brokers are employed, no third-party
reports are required to track sales proceeds.

A few studies (Clotfelter, 1983; Feinstein, 1991, among others) analyze TCMP
data to assess the determinants of overall tax evasion, and their findings guide our
tests. However, to our knowledge, Poterba (1987) is the sole study that attempts to
identify determinants of capital gains tax compliance. Using aggregated TCMP
data for 6 years from 1965 to 1982, he regresses the total capital gains reported on
tax returns as a percentage of the total IRS-computed capital gains on the marginal

¹In a review of 1979 tax returns, Thompson (1987) notes that capital gains tax compliance rates for
stocks, the asset under investigation in this study, is higher than for business property and personal
residences. Also, the 1988 increase in capital gains compliance may be attributable to the Tax Reform
Act of 1986 requirement that brokers report proceeds to the IRS.
tax rate applicable to capital gains. He reports an inverse relation, which is
consistent with capital gains tax compliance decreasing in marginal tax rates.
Limited to six observations, Poterba (1987) is unable to test for other deter-
minants.

Although it is virtually the sole source of capital gains tax compliance
information to date, there are at least three reasons why TCMP data are limited in
their ability to determine cross-taxpayer differences in capital gains tax. One,
desired information may be lost or obscured through the TCMP audit and
subsequent aggregation for the disclosure process. Two, the IRS audits may fail,
leaving noncompliance undetected or compliance mischaracterized. Three, TCMP
data do not report whether capital gains noncompliance is attributable to proceeds
understatement, basis overstatement, or both, and therefore are not useful for
differentiating noncompliance across third-party reporting requirements for
proceeds and bases.

This study provides a more comprehensive analysis of the determinants of
capital gains tax compliance for one major transaction. We match 1989 individual
tax returns provided by the IRS to detailed shareholder records provided by RJR
Nabisco for North Carolina shareholders who sold their common stock in 1989
during the company’s LBO. Capital gains tax compliance is estimated at the
individual shareholder level. We compare the proceeds from the sale of shares
during the LBO and the tax bases of those shares, as reported on the actual income
tax return, with the proceeds and tax bases, as estimated from RJR Nabisco’s
shareholder records. Discrepancies in the amount of reported net gain and net gain
per shareholder records are interpreted as noncompliance.

The RJR Nabisco LBO presents an intriguing setting to investigate capital gains
tax compliance. With a 79 percent takeover premium creating a then-record $25
billion price tag, the acquisition generated more capital gains than any previous
transaction. Moreover, the returns to noncompliance were substantial for many
shareholders examined in this study because they had held large positions for
many years in shares with low tax bases. On the other hand, concentrating on one
transaction limits the generalizability of the findings. Ideally, it would be
preferential to examine the determinants of capital gains tax compliance across
many transactions. Unfortunately, access to data comparable to those examined in
this study is unusual, and, thus, this paper is limited to this single, albeit
extraordinary, transaction.

We find that RJR Nabisco capital gains basis tax compliance is decreasing in
income including the RJR Nabisco capital gains, first-dollar marginal tax rates,
self-employment income, rental income, and interest expense classified as ‘other’
on the tax return. The income effect is consistent with Clotfelter (1983) and
The marginal tax rate finding confirms Poterba’s (1987) finding that the tax price
affects compliance. The low capital gains tax compliance by the self-employed
and renters (who are known to exploit opportunities for aggressive tax reporting)
and by taxpayers who report relatively large amounts in nebulous categories (such as other interest paid) is consistent with noncompliance being greatest among those taxpayers who perceive the penalty of detection to be lower.

The findings in this paper advance our understanding of the profiles of capital gains tax compliers and noncompliers and should interest tax administrators and policymakers. The IRS’s stated reason for permitting our access to confidential tax returns was a belief that these results could improve audit procedures and revenue estimates. The results also could affect ongoing debates concerning the appropriateness and level of capital gains taxation. Finally, to the extent this paper is useful in explaining cross-taxpayer variation in compliance, it extends our understanding of the incidence of the capital gains tax.

The remainder of the paper is organized as follows: Section 2 highlights relevant features of the RJR Nabisco LBO. Section 3 describes the data provided by the IRS and RJR Nabisco. Section 4 discusses the research design. Section 5 presents the empirical tests and results. Concluding remarks follow.

2. A model of tax evasion

Drawing on the model by Allingham and Sandmo (1972) and Yitzhaki (1974), we consider a taxpayer who chooses reported income, $R$, to maximize expected utility, $U$, expressed as

$$E(U) = (1 - p)U(I - \tau R) + pU(I - \tau R - L, \pi(I - R)), \tag{1}$$

where $p$ is the probability of detection, $I$ is true income, $\tau$ is the tax rate, $\pi$ is the evasion penalty, and $L$ is the individual-specific attenuation of the evasion penalty.

The interpretation of the penalty on detection differs slightly from that given by Allingham and Sandmo, who consider the penalty to be a linear proportion of unreported income with no individual taxpayer heterogeneity. Our model relaxes the assumptions of homogeneity and linearity. In practice, the penalty ranges from a minimum of zero (in the case where no additional penalty is charged and the taxpayer is only required to pay the tax on the unreported income) to imprisonment (in the case of tax fraud, which is a criminal act usually reserved for extreme cases of underreporting of income). IRS agents are given considerable leeway during an audit with respect to imposed penalties, some of which are dependent on the agent’s perceptions of the taxpayer’s intents and motives.

---

3Interest, which is also usually levied on unpaid tax liabilities, is ignored in the model. The general inferences of the model are unaffected by this assumption even to the extent that the taxpayer’s risk free rate on earnings are less than the rate charged by the IRS.
We posit that \( L_i \) is an (unspecified) nonlinear decreasing function of the amount of reported income. \( L_i \) is not a function of exogenous penalty rates or the probability of detection, although a taxpayer may incorporate those notions into his individual perceptions of \( L_i \). The essential notion of \( L_i \) is that two taxpayers may perceive the penalty on detection to be different, depending on their tax aggressiveness and that this perception is not necessarily linear. Although it is empirically difficult to estimate a taxpayer's perceived penalty on detection, it is probably safe to assume that aggressive taxpayers perceive the penalty to be lower. Although aggressiveness also is difficult to measure, our data permit us to construct proxies.

A first order necessary condition relating to (1) is

\[
\frac{dU}{dR} = -(1 - p)U'(I - \pi R) - \pi L_i (I - \pi R - L_i \pi (I - R)).
\]  (2)

The related second order necessary condition is

\[
\frac{d^2 U}{dR^2} = \pi^2 (1 - p)U''(I - \pi R) + \pi L_i (I - \pi R - L_i \pi (I - R)),
\]  (3)

which is strictly positive. To estimate the effect of a change in the perceived penalty on the amount of reported income, the first order condition is differentiated with respect to \( L_i \) and set equal to zero:

\[
\frac{\partial R}{\partial L_i} = -\frac{1}{d^2 U/dR^2} (\pi L_i (I - \pi R - L_i \pi (I - R)))
\]

\[
= -\pi^2 \frac{1}{d^2 U/dR^2} U'(I - \pi R - L_i \pi (I - R)),
\]  (4)

which is strictly positive, suggesting that when holding the level of the penalty, \( \pi \), constant, the level of reported income is increasing in \( L_i \).

We construct a proxy for \( L_i \) using three categories of income and expenses that prior research has found to be particularly subject to high levels of noncompliance. The first category consists of those items that are business activities where income and expense items are netted and few, if any, items are subject to third-party reporting requirements. An example would be Schedule C for sole proprietors. The second category consists of items described on the tax return as 'other', such as other income and other interest paid. These items are typically defined nebulously and usually not reported by third parties. This enables aggressive taxpayers to deduct otherwise nondeductible expenses (e.g., by classifying personal interest expense as other interest paid) or to avoid the self-employment tax.

\[\text{\footnotesize A third-party reporting requirement compels a payer or payee to report the amount of payments to the IRS. Examples include wages on form W-2, interest income on form 1099 and interest expense on form 1098.}\]
tax (e.g., by reporting self-employment income as other income thus avoiding
schedule SE and its self-employment tax). The third category includes cash and
non-cash charitable deductions, which in 1989 were not subject to third-party or
self-reporting requirements.

These three categories comprise our proxy for taxpayer aggressiveness in that
they represent opportunities for aggressive behavior, which may suggest a lower
perceived penalty on detection. It is possible that non-aggressive taxpayers find
themselves with income and deductions in one or more of these three categories as
well. The presence of this type of taxpayer would presumably add noise to our
measures but should not introduce a systematic bias.

A second way that aggressiveness may manifest itself in the model is that
aggressive taxpayers may assess the probability of detection as small. However,
this is unlikely because aggressive behavior on tax returns increases the probability
of detection. That is, the IRS is more likely to audit tax returns that contain
abnormally high levels of aggressive categories, a selection process not unknown
to taxpayers. 5

3. RJR Nabisco’s leveraged buyout

On February 9, 1989, Kohlberg, Kravis, Roberts & Co completed a $25 billion
tender offer of RJR Nabisco outstanding shares. The acquisition ended a struggle
for the company that had begun with a management-led bid on October 20, 1988.
The acquisition was unprecedented in its magnitude, four-fold the next largest
purchase to date.

RJR Nabisco’s LBO potentially provides an unusually powerful setting to
analyze cross-taxpayer variation in capital gains tax compliance because it should
have magnified shareholders’ tendencies to underreport for at least five reasons.
First, every stock sale examined in this study triggered capital gains. The share
price jumped from $55.875 to a then-record $77.25 with announcement of the first
takeover bid (October 20, 1988), finally peaking at $100 on the last trading day
(February 8, 1989). During 1989, the tax year investigated in this study, the price
always exceeded $90.

Second, the capital gains were substantial, making returns to evasion high. Not
only did the stock sell at a substantial premium, but the tax bases of the selling
shareholders examined in this study were unusually low. Examining the same RJR
Nabisco shareholder data that this study uses, Landsman and Shackelford (1995)
estimate that selling individual and corporate shareholders had a mean tax basis of
$21 per share. Furthermore, they acknowledge that their computations likely
overstate basis because they assume tax basis is adjusted whenever shares are

---

5 Although exact IRS thresholds for audit are unknown, most taxpayers are aware that certain
categories of income or expense are subject to a higher percentage of audit. Even popular tax
preparation software packages (e.g., Turbo Tax) contain ‘red flags’ for these types of income.
issued, a condition that does not hold for inter vivos gifts. Thus, at a minimum, the
capital gains exceeded 77 percent of proceeds for the average shareholder
throughout 1989. Alternatively stated, the capital gains taxes for the typical
shareholder examined in this study exceeded 25 percent of proceeds, assuming a
33 percent combined Federal and North Carolina capital gains tax rate.\footnote{The 33 percent equals a maximum Federal capital gains tax rate of 28 percent plus a deductible maximum North Carolina capital gains tax rate of 7 percent (28 + 7(1 − 0.28)).}

Third, the investors examined in this study had substantial RJR Nabisco
holdings. Landsman and Shackelford (1995) report that the mean (median) taxable
shareholder examined in this database held 1301 (370) shares, ranging from one
share to 875,000 shares. The mean holding was worth more than $117,000
throughout 1989.

Fourth, compliance likely was undermined by inadequate record keeping,
prompted by the lengthy holding period for most shares. Landsman and
Shackelford (1995) estimate that the mean holding period for North Carolina
shareholders was 9.9 years, compared with 1.2 for other RJR Nabisco sharehol-
ders. Many shares had been held for decades, which they attribute to an extensive
employee stock program and unusual shareholder loyalty in North Carolina. Tax
professionals tell us that many shareholders had no records of the price originally
paid for the stock, in part because many shares had been gifted within families,
and tax bases were estimated for many returns. The lack of evidence establishing
tax bases likely extenuated any proclivities to understate capital gains.

Finally, our private discussions with tax compliance officials in North Carolina
suggest that the political clout of the company and the business and political
prominence in the state of some of its leading shareholders may have diminished
the aggressiveness of audits by state tax officials. If so, underreporting (at least at
the state level) may have been unusually large for this transaction.

On the other hand, the visibility of the RJR Nabisco LBO may have dampened
evasion, reducing the power of these tests to distinguish compliers and noncom-
pliers. The fact that a company of RJR Nabisco’s size could be acquired through
an LBO shocked the investment community and attracted extensive press
attention. Despite the large gains to noncompliance, if noncompliers believed that
the taxing authorities (at least at the Federal level) would increase their scrutiny
accordingly, they may have been more compliant than usual.

4. Data

4.1. RJR Nabisco shareholder records

RJR Nabisco gave us detailed records for its shareholders with North Carolina
The records include each shareholder’s name, address, tax identification number,
the number of shares issued and the dates issued, and the number of shares sold
and the dates sold. The records correctly report each investor's total RJR Nabisco
holdings with one exception. The identity of investors' holding shares in 'street
name,' i.e., shareholdings with investment advisory services, is unknown to the
company and, consequently, unavailable to us.

Landsman and Shackelford (1995) detail reconstruction of the sample to
approximate the taxable shareholders on October 19, 1988, the day before the first
offer to buy the company. The sample (hereafter, the RJR Nabisco sample)
includes 18,918 individuals and corporations holding 23 million shares.

4.2. Individual tax returns

The IRS gave us 1989 Federal tax returns for 820 individuals who sold their
RJR Nabisco stock between January 1, 1989 and February 8, 1989 (hereafter, the
IRS sample). The authors have an assignment agreement under the Intergov-
ernmental Personnel Act of 1970 between their employer and the IRS District Office
Research and Analysis Division in Greensboro, North Carolina. The agreement
permits them access to actual tax return data. They are subject to the same
confidentiality requirements that bind IRS personnel. The analysis is limited to
1989 sales because the IRS had destroyed 1988 tax returns when the study was
started.

We requested tax returns for 1000 individual shareholders, selected from the
RJR Nabisco sample. Half the shareholders were selected through dollar-value
sampling and half through random sampling. These procedures were designed to
overweight the sample toward the largest shareholders and still include substantial
portions of the entire population. Of the 1000 shareholders requested, the IRS was
unable to find the tax returns for 151. The potential sample was further reduced by
29 for spouses who filed joint returns with other shareholders.6 We input selected
data from the personal tax return (Form 1040 and its schedules) for the remaining
820 tax returns. Internal validity was assessed by matching subtotals, such as total
income and taxable income, against their components.

We also collected as much detail as possible from schedule D (Capital Gains
and Losses) about the RJR Nabisco sales, including the number of shares sold, the
dates sold and acquired, and the proceeds and the tax bases. Unfortunately, the
details were often scant. Few taxpayers recorded each capital transaction in detail,
as requested by the Schedule D form. Taxpayers often provided only an
aggregated total for all capital gains. Sometimes the sole information supporting
the capital gains amount was a single line in the return, stating 'various' or similar

6The RJR Nabisco sample does not indicate married couples and whether they file joint returns. In
our tests, the RJR Nabisco sample data for spouses are aggregated and compared with the single joint
return.
nomenclature for the name of the assets sold, the number of units sold, the dates sold and acquired, and the proceeds and tax bases.

4.3. Matching RJR Nabisco and IRS samples

Ideally, we would match the RJR Nabisco and IRS samples on a transaction-by-transaction basis. Unfortunately, the inadequate tax return disclosures limit our ability to match RJR Nabisco and IRS sample data in this detailed fashion for all shareholder-taxpayers. Instead, we resort to matching the samples on total shares and total proceeds. Consequently, tests are conducted only on individuals for whom total RJR Nabisco proceeds as reported on the tax return divided by the proceeds per the shareholder records varies from one by more than three percent. That is, we discard observations for which the ratio of the two proceeds is less than 97 percent or greater than 103 percent, reducing the sample from 820 observations to 411 observations. Among other casualties of this screen are individuals who did not separately detail in their tax returns the proceeds received from RJR Nabisco transactions.

This screening procedure is necessary because of the large number of shares held in street name by RJR Nabisco shareholders and it is not possible for us to allocate general brokerage accounts among taxpayers. Without adopting this convention, it is possible that the individual’s tax return would show higher proceeds than the RJR Nabisco records as the RJR Nabisco records would not reveal which shares were held in street name. Thus, to properly measure basis overstatement, it is crucial that we compare samples matched on proceeds. By using a matching threshold of 3 percent, which is a lower bound on brokerage fees, we remove the effects of street name holdings.\(^7\) Note that by symmetrically removing observations for which proceeds do not match above and below a threshold also removes taxpayers who underreport proceeds, and therefore focuses the study on a specific type of capital gains noncompliance, namely basis underreporting.\(^8\)

Besides evasion and unintentional errors, the total shares and proceeds per the RJR Nabisco sample may not match the comparable figures for the IRS sample for two additional reasons. First, the RJR Nabisco records do not indicate whether a taxpayer disposed of his shares in 1989 in a taxable transaction. If shareholders transferred ownership through inter vivos gifts, the actual proceeds in the tax return will not match the assumed proceeds from the RJR Nabisco records. For example, if a shareholder gave stock to his daughter, the RJR Nabisco sample

\(^7\)Results are robust to a matching threshold of 6 percent.

\(^8\)Given that street name holdings would never cause the tax return proceeds to be higher than the RJR Nabisco records, we considered removing only those observations where the proceeds on the tax return were higher than RJR Nabisco. The problem with this method is that it may introduce bias and is less parsimonious than symmetrically trimming the sample.
would show a disposal, which we would assume is a capital gain, but his tax return
would correctly show no capital gain. Unfortunately, the data do not enable us to
reconcile proceeds for donors. Similar differences between the RJR Nabisco and
IRS samples could arise from 1989 charitable contributions of stock. However, to
the extent taxpayers itemize charitable contributions and detail the contribution of
RJR Nabisco stock, shares contributed to charities could be identified from
Schedule A (Itemized Deductions) of the individual's tax return. Accordingly, we
find that 20 individuals in this study itemized charitable contributions of nearly
$3.5 million of RJR Nabisco stock. For these 20 shareholders, the RJR Nabisco
and IRS samples are reconciled by adjusting shares and proceeds for the charitable
gifts.

Second, RJR Nabisco shares split on January 31, 1929, April 9, 1959,
measurement error in our estimate of tax basis using shareholder records. When
shares split, the tax basis of the original shares is spread pro-rata across the
original shares and the new shares. We assign shares issued from splits a basis of
zero. This obviates the need to assign basis among shares after a split. For
example, if a shareholder had one share with a basis of $10 before a two-for-one
split, the $10 of basis should be split evenly between the original share and the
share received giving each share a $5 basis. Our method would assign a zero basis
to the split share while keeping the basis of the original share at $10. Either
method will provide for a correct accounting of basis to the extent that all shares
(or original and split) are sold in the RJR Nabisco acquisition. This method does not
require us to allocate fractional shares multiple times across splits (i.e., when split
shares become split themselves), while maintaining a proper allocation of basis
across total shares.

5. Research design

5.1. Regression equation

Motivated by the theoretical model in Section 2, Eq. (5) provides the primary
empirical model for estimating the determinants of capital gains tax compliance:

\[
\text{Ratio} = a_0 + a_1 \text{Wages} + a_2 \text{Taxable Interest} + a_3 \text{Exempt Interest} \\
+ a_4 \text{Dividends} + a_5 \text{Pension} + a_6 \text{RJR Gain} + a_7 \text{MTR} \\
+ a_8 \text{Homeownership} + a_9 \text{Age} + a_{10} \text{Single} + a_{11} \text{Male} + a_{12} \text{Index} \\
+ a_{13} \text{Table} + a_{14} \text{Index} \times \text{Table} + a_{15} \text{Preparer} + a_{16} \text{Sch.C} \\
+ a_{17} \text{Sch.F} + a_{18} \text{Rent} + a_{19} \text{Other Income} + a_{20} \text{Other Taxes} \\
+ a_{21} \text{Other Interest Paid} + a_{22} \text{Cash Contributions} \\
+ a_{23} \text{Noncash Contributions} + \epsilon.
\] (5)
To control for the effects of heteroskedasticity, we deflate all monetary explanatory variables by total income per the tax return, item 23 on the form 1040 (i.e., income before the adjustments that create AGI), with one adjustment. The income adjustment is the substitution of RJR Nabisco capital gains, as computed using the RJR Nabisco shareholder records (RJR gain), for the capital gains reported on the tax return (IRS gain). The purpose of the substitution is to create a scalar that is unaffected by the potential RJR Nabisco capital gains tax evasion.9

Even after scaling, the White (1980) test for homoscedasticity remains rejected at conventional levels. As a result, all results are presented using asymptotically corrected standard errors. To control for the effects of outliers, we employ a method suggested by Belsley et al. (1980) and delete all observations with a Cook’s D statistic greater than one. This eliminates five observations, leaving a final sample of 406 taxpayers.10 Table 1, Panel A, provides descriptive statistics of selected tax return items and the capital gains computed using RJR Nabisco shareholder records for this sample. Table 1, Panel B, provides similar summary statistics for the regression variables.

Ratio, the dependent variable, is designed to measure a taxpayer’s RJR Nabisco capital gains tax compliance. Following the model in Section 2, the dependent variable is the reported RJR Nabisco capital gains per the tax return (IRS gain). Using the conventional approach of scaling both sides of the equation by the same factor to address heteroscedasticity, IRS gain should then be scaled by total income. Instead, for reasons detailed below, we scale the IRS gain by the estimated RJR Nabisco capital gains per the company’s shareholder records (RJR gain). That is, Ratio is the IRS gain/RJR gain and is increasing in compliance. The proceeds (tax bases) needed to estimate the RJR gain are the product of the number of shares sold (issued), as disclosed in the shareholder records, and the closing price of the stock on the date of sale (purchase), adjusted for stock splits.11 Table 1, Panel B, indicates the mean (median) Ratio is 1.1 (1.0) for the full sample of 406 observations.

The reason for scaling by the RJR gain, rather than total income, is that total income would introduce unwanted variation in the regression. For example,
Table 1
Descriptive statistics

<table>
<thead>
<tr>
<th>Variable name (description)</th>
<th>Full Sample (n = 406)</th>
<th>Ratio &lt; 1.3 (n = 352)</th>
<th>Ratio &lt; 1.0 (n = 181)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Med.</td>
<td>S.D.</td>
</tr>
<tr>
<td>Wages</td>
<td>25.3</td>
<td>0.0</td>
<td>144.5</td>
</tr>
<tr>
<td>Taxable Interest</td>
<td>32.7</td>
<td>15.9</td>
<td>57.3</td>
</tr>
<tr>
<td>Exempt Interest</td>
<td>9.2</td>
<td>0.0</td>
<td>34.5</td>
</tr>
<tr>
<td>Dividends</td>
<td>21.2</td>
<td>5.6</td>
<td>44.4</td>
</tr>
<tr>
<td>Pension (includes IRA income)</td>
<td>8.2</td>
<td>0.3</td>
<td>25.8</td>
</tr>
<tr>
<td>JRJ Gain</td>
<td>418.9</td>
<td>214.1</td>
<td>782.7</td>
</tr>
<tr>
<td>Sch C (net business income)</td>
<td>1.4</td>
<td>0.0</td>
<td>12.0</td>
</tr>
<tr>
<td>Sch F (net farm income)</td>
<td>-0.3</td>
<td>0.0</td>
<td>2.8</td>
</tr>
<tr>
<td>Rent (net rental income on Sch E)</td>
<td>1.2</td>
<td>0.0</td>
<td>8.6</td>
</tr>
<tr>
<td>Other Income (form 1040, line 22)</td>
<td>0.4</td>
<td>0.0</td>
<td>2.9</td>
</tr>
<tr>
<td>Other Taxes (Sch A, line 7)</td>
<td>0.9</td>
<td>0.0</td>
<td>1.6</td>
</tr>
<tr>
<td>Other Interest Paid (Sch A, line 12b)</td>
<td>0.1</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>Cash Contributions (Sch A, line 14)</td>
<td>4.2</td>
<td>1.3</td>
<td>9.9</td>
</tr>
<tr>
<td>Noncash Contributions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Sch A, line 15)</td>
<td>3.1</td>
<td>0.0</td>
<td>17.8</td>
</tr>
</tbody>
</table>

(B) Descriptive statistics for regression variables. All monetary explanatory variables are deflated by total income.

Dependent variables:
- Ratio (IRS gain/JRJ gain) 1.145 1.008 0.974 0.997 0.143 0.895 0.952 0.137

Explanatory variables:
- Wages 0.132 0.000 0.260 0.117 0.000 0.243 0.136 0.000 0.257
- Taxable Interest 0.019 0.001 0.113 0.009 0.000 0.115 0.009 0.000 0.106
- Exempt Interest 0.014 0.000 0.049 0.014 0.000 0.050 0.014 0.000 0.059
- Dividends 0.003 0.020 0.092 0.051 0.021 0.078 0.051 0.026 0.071
- Pension (includes IRA income) 0.049 0.000 0.123 0.042 0.001 0.105 0.051 0.002 0.110
- JRJ Gain 0.632 0.772 0.305 0.657 0.700 0.294 0.618 0.754 0.309
<p>| 484 | Tax Price: MTR (first-dollar marginal tax rate) | 0.204 | 0.150 | 0.109 | 0.204 | 0.150 | 0.110 | 0.210 | 0.215 | 0.105 |
| 486 | Socioeconomic characteristics | | | | | | | | | |
| 487 | Homeownership (1 if report residential) | 0.727 | 1.000 | 0.446 | 0.729 | 1.000 | 0.445 | 0.700 | 1.000 | 0.459 |
| 489 | Age (1 if positive pension) | 0.411 | 0.000 | 0.493 | 0.425 | 0.000 | 0.495 | 0.447 | 0.000 | 0.499 |
| 492 | Security income; else 0 | 0.414 | 0.000 | 0.493 | 0.428 | 0.000 | 0.495 | 0.458 | 0.000 | 0.500 |
| 493 | Single (1 is not married) | 0.073 | 0.000 | 0.260 | 0.069 | 0.000 | 0.254 | 0.063 | 0.000 | 0.244 |
| 494 | Filing joint return; else 0 | 0.073 | 0.000 | 0.260 | 0.069 | 0.000 | 0.254 | 0.063 | 0.000 | 0.244 |
| 497 | Table (1 if compute tax) | | | | | | | | | |
| 499 | Paper with tax table; else 0 | 0.195 | 0.000 | 0.396 | 0.182 | 0.000 | 0.387 | 0.216 | 0.000 | 0.412 |
| 501 | Index*Table | 4.540 | 0.000 | 11.089 | 4.157 | 0.000 | 10.641 | 4.800 | 0.000 | 11.341 |
| 502 | Preparer (1 if used paid) | | | | | | | | | |
| 503 | tax preparer; else 0 | 0.859 | 1.000 | 0.349 | 0.865 | 1.000 | 0.343 | 0.842 | 1.000 | 0.366 |</p>
<table>
<thead>
<tr>
<th>Variable name</th>
<th>Full Sample ($n = 406$)</th>
<th>Ratio &lt; 1.3 ($n = 352$)</th>
<th>Ratio &lt; 1.0 ($n = 181$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggressive opportunities measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sch C (net business income)</td>
<td>0.009</td>
<td>0.000</td>
<td>0.062</td>
</tr>
<tr>
<td>Sch F (net farm income)</td>
<td>-0.010</td>
<td>0.000</td>
<td>0.158</td>
</tr>
<tr>
<td>Rent (net rental income on Sch E)</td>
<td>0.005</td>
<td>0.000</td>
<td>0.045</td>
</tr>
<tr>
<td>Other Income (form 1040, line 22)</td>
<td>0.001</td>
<td>0.000</td>
<td>0.021</td>
</tr>
<tr>
<td>Other Taxes (Sch A, line 7)</td>
<td>0.002</td>
<td>0.001</td>
<td>0.003</td>
</tr>
<tr>
<td>Other Interest Paid (Sch A, line 13b)</td>
<td>0.001</td>
<td>0.000</td>
<td>0.003</td>
</tr>
<tr>
<td>Cash Contributions (Sch A, line 14)</td>
<td>0.012</td>
<td>0.004</td>
<td>0.022</td>
</tr>
<tr>
<td>Noncash Contributions (Sch A, line 15)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.003</td>
<td>0.000</td>
<td>0.020</td>
</tr>
</tbody>
</table>

*The scalar is total income before the adjustments that create adjusted gross income (line 23 on the taxpayer's form 1040) less the reported capital gains on the sale of RJR stock plus the capital gains on the sale, as computed using RJR Nabisco shareholder records.*
suppose two shareholders are identical in all regards (including capital gains tax
compliance), except one receives wages. If we scale the dependent variable by
total income, the shareholder enjoying wages has a smaller dependent variable
than the non-wage shareholder does. To control for this wage variation, a control
variable (e.g., wages) would need to be added to the explanatory variables.
However, wages are already included as a dependent variable because we seek to
determine whether capital gains compliance varies across income streams, such as
wages. By scaling the dependent variable by the RJR gain, rather than total
income, the variation in the dependent variable captures the variation across
shareholders in capital gains compliance while avoiding other sources of income
variation across shareholders. In other words, scaling by the RJR gain enables the
coefficients to be interpreted as the marginal impact of the explanatory variable on
noncompliance, after controlling for variation in income.

5.2. Explanatory variables

The explanatory variables can be segregated into income streams (e.g., wages,
dividends, interest, etc.), a tax price (marginal tax rates), socioeconomic charac-
teristics (evaluated in previous compliance studies or analyzable in this study
because of access to a different data source), and aggressive avoidance oppor-
tunities that are intended to capture the perceived penalty of detection discussed in
Section 2.

5.2.1. Income

As demonstrated in Allingham and Sandmo (1972), the relation between
compliance and income depends on assumptions about the taxpayer’s relative risk
aversion. Reported income as a fraction of true income increases, remains
constant, or decreases as relative risk aversion is an increasing, constant, or
decreasing function of income. Although absolute risk aversion is generally
presumed to be decreasing in wealth, it is difficult to predict the effects of relative
risk aversion. Consistent with expectations of decreasing absolute risk aversion,
Clotfelter (1983) finds that overall compliance is decreasing in his proxy for
wealth (AGI less Federal taxes) in his examination of the 1969 TCMP survey
results.12 Similarly, Feinstein (1991) reports compliance decreasing in wealth, as
measured by AGI, when 1982 and 1985 TCMP data are analyzed separately.

12As an aside, Clotfelter’s (1983, p. 368) description of wealth effects states “[T]hese results are
consistent with the hypothesis of increasing absolute risk aversion, commonly accepted as a reasonable
assumption in models of individual choice under uncertainty.” In our private discussions with Professor
Clotfelter, he indicated to us that this statement is erroneous. Instead, the results are consistent with
decreasing absolute risk aversion, which is our characterization of his results in this paper.
However, when he pools the 2 years, Feinstein (1991) finds no evidence that
wealth affects overall compliance.

The regression equation includes five income flows: wages, taxable interest
income, tax-exempt interest income, dividend income, and pension income. It also
includes the amount of gain related to the RJR Nabisco transaction (as computed
using RJR Nabisco shareholder records) to measure the relation between the size
of the gain and compliance.13 Returns to financial capital that may involve tax
shelters, such as rents and farm income, are excluded from this wealth measure.

Table 1, Panel B, shows that the RJR Nabisco capital gains comprise most of
the shareholders’ income in 1989, representing 63 percent of total income (where
capital gains per RJR Nabisco shareholder records substitutes for reported capital
gains), on average, in the full sample. Wages are an unimportant income source for
most of these taxpayers. Table 1, Panel A, shows mean wages of $25 thousand or
4 percent of AGI. Median wages are zero. Excluding the RJR Nabisco capital
gains, mean wages still constitute only 20 percent of their taxpayers’ AGI.

5.2.2. Tax price

As shown in Allingham and Sandmo (1972) and Yitzhaki (1974), the relation
between tax rates and noncompliance varies with risk aversion and the nature of
the penalty. If an individual’s absolute risk aversion is decreasing in income and
the penalty is levied on the amount of tax evaded, compliance is increasing in tax
rates arising from an income effect and the absence of a substitution effect. If the
fine is levied on unreported income as opposed to the underpaid tax, the results are
less clear. In a similar fashion, our model is unable to predict the relation between
tax rates and noncompliance. As noted above, Poterba (1987) reports capital gains
tax compliance decreases in marginal tax rates using TCMP data for 6 years from
1965 to 1982. Similarly, Clotfelter (1983) and Feinstein (1991), when he pools
1982 and 1985, finds that overall tax compliance is decreasing in marginal tax
rates.

We consider two measures of the marginal tax rate, a first-dollar rate and a
last-dollar rate. The first-dollar rate is the amount of additional tax levied on the
first dollar of reported RJR Nabisco capital gains. The last-dollar rate is the
additional tax on the last dollar of reported gain. Unfortunately, the last-dollar
marginal tax rates of the investors in this study are sufficiently similar to weaken
tests for a relation between compliance and last-dollar marginal tax rates. The
large RJR Nabisco capital gain combined with substantial other income leaves 92
percent of our taxpayers in the maximum statutory tax bracket of 28 percent for
the RJR Nabisco capital gains. We use an indicator variable in an additional set of
regressions to distinguish the shareholders in the 28 percent tax bracket from
shareholders in lower tax brackets. Unreported results fail to find any relation

13We use the gain as reported by RJR Nabisco and not the taxpayer to mitigate possible endogeneity
bias.
between the last-dollar rate and compliance, although all other inferences were unaltered. This result is likely attributable to the lack of variation in the rate. The remaining analysis is conducted by including the first-dollar tax rate (MTR).

5.2.3. Socioeconomic characteristics

5.2.3.1. Homeownership

The third set of explanatory variables addresses socioeconomic characteristics, beginning with homeownership. To our knowledge, prior studies have not evaluated the impact of homeownership on compliance. Although the model in Section 2 cannot be used directly to develop a prediction for the relation between homeownership and noncompliance, homeownership can induce risk aversion if the homeowner perceives the potential loss of underreporting to be greater than other taxpayers. If the IRS discovers noncompliance and the taxpayer does not have sufficient funds to pay the tax and penalty, the IRS may take the home as payment. If taxpayers perceive the cost of losing a home as particularly high, noncompliance may be decreasing in homeownership. However, left with no clear prediction, the sign of the relation is an empirical issue. Homeownership is measured by an indicator variable, Homeownership, which equals one for the 73 percent of sample taxpayers who itemize real estate taxes, and zero otherwise. Because 94 percent of our sample itemizes, the possibility that homeowners are overlooked by using the itemized deductions from Schedule A is not a major concern. 14

5.2.3.2. Age

The model in Section 2 also provides no direct prediction regarding the relation between age and noncompliance. Prior studies generally have shown that the elderly evade less than younger taxpayers do. In his analysis of the 1969 TCMP survey, Clotfelter (1983) finds that taxpayers who are at least 65 years old are less likely to underreport total taxable income. Similarly, using pooled 1982 and 1985 TCMP data and evaluating 1982 separately, Feinstein (1991) reports that underreporting is less for 65 or older taxpayers. No relation is detected when 1985 is estimated separately.

Although there is no direct measure of a taxpayer’s age on the tax return, there are several constructive proxies, including pension income, IRA income, or Social Security income. We construct an indicator variable (Age) that equals one if the taxpayer received income from any of these three sources, zero otherwise.

5.2.3.3. Marital status

As with age, the model provides no direct prediction about the relation between

---

14 Itemizing was common for these taxpayers partly because North Carolina taxes generated by RJR Nabisco sales were deductible in 1989.
marital status and noncompliance. Prior studies generally have shown that married
couples underreport total income more than other taxpayers do. In his analysis of
the 1969 TCMP survey, Clotfelter (1983) finds that married taxpayers understate
more unless they file Schedule C (indicating self-employment) or a Schedule F
(indicating farming activities) in which case no difference is noted between single
that underreporting does not vary with marital status. However, when each year is
estimated separately, couples underreport more than singles. We include an
indicator variable, \textit{Single}, equaling one for all taxpayers, other than couples filing
jointly. Table 1, Panel B, indicates 41 percent of the full sample do not file joint
returns.

5.2.3.4. Gender

We further dichotomize taxpayers who do not file joint returns by gender. To
our knowledge, prior empirical studies have ignored any possible gender differences
in compliance and we posit no directional prediction. \textit{Male} is an indicator
variable, equaling one for all taxpayers, other than those filing joint returns, who
are male. Among the non-joint return filers, women outnumber men nearly
six-to-one. As with other taxpayer characteristics, the model does not provide a
direct prediction about the relation between gender and noncompliance. Evaluating
three of the socioeconomic measures (age, marital status and gender) jointly, we
surmise that the sample is highly weighted toward elderly widows.

5.2.3.5. Evasion

Slemrod (1985) develops a measure of tax evasion based on the final two digits
of taxable income. Most taxpayers determine their tax liability using tax tables that
provide the same dollar amount for taxable income within \$50 ranges. For
example, the total tax liability is identical for all taxpayers with taxable income
between \$18,000 and \$18,049. Slemrod (1985) presents evidence suggesting that
evaders are more likely to report taxable income in the upper portion of the range.

To test for an association between tax table manipulation and RJR Nabisco
capital gains compliance, we include an evasion index in our regression analysis.
\textit{Index} is the final two digits of taxable income, less 50 for values exceeding 49. It
averages 25 in this study. Because the taxable income of 81 percent of our sample
exceeds the limit for using the tax tables, a categorical variable, \textit{Table}, is required
to distinguish filers using the tax tables (and consequently able to lower taxes
using Slemrod's method) from other taxpayers (who cannot avail themselves of
this evasion technique). We predict that the interaction of the two variables,
\textit{Index}*\textit{Table}, will be negative, indicating an association between reporting in the
upper range of the tax table and underreporting RJR Nabisco capital gains.

5.2.3.6. Paid tax preparer

\textit{Preparer} is a categorical variable indicating whether a tax professional was paid
to prepare the tax return. The directional effect of a tax preparer is uncertain. On
the one hand, a professional faces penalties, including criminal charges, for completing a fraudulent tax return, implying a positive coefficient for Preparer.

On the other hand, a tax preparer interprets the law in his client’s favor and likely detects errors that favor the government. Thus, a negative coefficient for Preparer could occur because a professional reduces the likelihood of overstating capital gains.

Aggregate statistics from the 1979 TCMP indicates little variation in voluntary compliance rates across various paid tax preparer categories and self-prepared returns. Klepper et al. (1991) find that tax preparers lead to higher compliance with relatively unambiguous tax return line items, such as interest, and lower compliance with more ambiguous line items, such as business income. Correct calculation of capital gain proceeds is straightforward. Basis calculations, however, can be complicated by gifts, stock splits, and general record keeping deficiencies. Because the model makes no prediction as to the relation between use of a preparer and compliance, we include Preparer without any coefficient sign prediction.

5.2.4. Perceived penalty on noncompliance, \( L_i \)

Prior studies have included additional measures for taxpayer activities where evasion is considered most likely. For example, L’Einstien (1991) finds that underreporting is greater for taxpayers who report self-employment income and farm income, which he attributes to complex tax laws that create opportunities to conceal income. Similarly, Slemrod (1985) characterizes taxpayers with adjustments to income, itemized deductions, self-employment income, partnership income, small business income, estate income (pass through income to a beneficiary from a decedent’s estate until such a time as the estate’s assets are distributed) or trust income as enjoying ‘fungible’ income and deductions and more likely to evade.\(^{15}\) Our study uses these variables as proxies for a taxpayer’s perception of the penalty on revealed noncompliance, creating eight separate measures of \( L_i \).

The first group contains the amount of net business activities as reported on Schedule C (for 49 self-employed taxpayers), Schedule F (24 Farming activities), and Schedule E (111 Rental activities). Table 1, Panel B, indicates self-employment income averages almost 0.9 percent of gross income in the full sample. Rental income is 0.5 percent of gross income. On average, the farmers in the sample report net losses, consistent with farming being a tax shelter activity for these taxpayers. These forms require the taxpayer to collect the revenues and

\(^{15}\) The IRS produces its own internally generated measure of overall expected tax compliance based on a variety of taxpayer characteristics. Unfortunately, the tax files (i.e., the tax returns and internal communications within the IRS) that we examined only contained this measure, known as the DIF score, for 12 sample taxpayers. Hence, though determining whether overall tax compliance as measured by the IRS is a useful predictor of RJR Nabisco capital gains tax compliance would have been interesting, we do not have enough data to conduct this test.
expenses related to business activities and present them in aggregated form. This
aggregation, along with the lack of third-party reporting for most of the income
and expense items, provides an aggressive taxpayer with opportunity for noncom-
pliance.

The second group consists of items described on the tax return as ‘other’ and
includes other income, other tax deductions and other interest deductions. Table 1,
Panel B, indicates that other taxes average 0.2 percent of total income. The other
two variables average half this amount. These items are a nebulous aggregation of
miscellaneous amounts generally not subject to third-party reporting and therefore
provide an additional opportunity for aggressive taxpayer behavior.

The last group is charitable deductions, both cash and non-cash. Taxpayers have
long been suspected of overstating cash (Cash Contributions) and non-cash
charitable contributions (Noncash Contributions). Accordingly, documentation is
now required to claim charitable contributions, although the documentation
requirements were substantially less rigorous during the investigation period and
thus provided an additional opportunity for noncompliance. Table 1, Panel B,
indicates that the full sample’s cash contributions average 1.2 percent of total
income, four times the size of non-cash giving.

5.2.5. Alternative regressors

Access to the complete tax returns of the sample shareholders obviously
provides us with many possible explanatory variables, most of which are not
expressly suggested by the model. Alternative regressors were excluded from the
analysis for two reasons. First, they offer insufficient variation across taxpayers.
Specifically, taxable refunds, alimony received and paid, alternative minimum
taxes, tip taxes, points paid on mortgage loans, other mortgage interest paid but not
reported on a 1099, charitable contribution carryovers, casualty losses, deductible
medical expenses and moving expenses are omitted because less than 10 taxpayers
recorded positive values for these items. Second, another set of alternative
regressors was judged to be subsumed by superior measures of the underlying
construct and did not affect the inferences when included in the regression. Gains
on the sale of business property, self-employment insurance deductions, self-
employment taxes, and Keogh deductions were considered subsumed by business
income. State and local tax deductions are captured by wealth. The number of
exemptions is subsumed by taxpayer age, i.e. elderly seldom have dependent
children. Finally, estate income was judged captured by age and wealth.

5.3. Measurement error and subsamples

The regressions in this study are conducted using the full sample of 406
taxpayers and two subsamples, whose descriptive statistics are detailed in the
second and third columns of Table 1. Subsamples are employed to address
concerns related to potential measurement error that may exist in the tax bases of
shares received from gifts. As noted above, tax bases are estimated assuming the
price of the stock on the issue date establishes the tax basis per share. This
assumption is only valid for stock acquired through purchase or inheritance. As
noted above, with inter vivos gifts, the donee’s tax basis in the stock is the same as
the donor’s tax basis. For example, if a mother gives shares to her son, his tax
basis in the shares after the gift is the same as his mother’s tax basis in the shares
before the gift. This carryover of basis cannot be detected using shareholder
records, which simply disclose that shares were issued to the son.

Consequently, to the extent taxpayers sold RJR Nabisco shares they received
through gifts, our estimates of the RJR gain likely are understated because the
share price rose steadily for many years. Unfortunately, gifted stock likely is
nontrivial in this study. Many stockholders in the RJR Nabisco database share the
same surname and address. Landsman and Shackelford (1995) report anecdotes
about longtime shareholder loyalty and concentrated ownership within North
Carolina families, increasing the probability that inter vivos giving was widespread
among sample taxpayers. (Note that similar problems do not arise with inherited
shares because tax basis is adjusted at death to the current market price.)

Our inability to identify shares transferred via gifts introduces measurement
error into the estimation procedures. Thus, we supplement our analysis with two
subsamples. The first subsample includes the 352 observations where \( \text{Ratio} \) is less
than its mean value of 1.3. The second subsample includes the 181 observations
where \( \text{Ratio} \) is not greater than one.

By removing observations that report more gain on the tax return than on the
RJR Nabisco records, the truncation in the second subsample likely eliminates
most measurement error attributable to gifts. For example, for the non-truncated
sample of 406 taxpayers, the mean RJR Nabisco capital gains per the IRS sample,
i.e. the IRS gain, is $455, compared with a mean RJR Nabisco capital gains per
the RJR Nabisco sample, i.e. the RJR gain, of $419, as shown in Table 1, Panel A.
This apparent (and obviously misleading) overreporting of gain on the tax returns
(more than 8 percent of RJR gains) is likely attributable to mismeasurement
associated with gifts. Truncation of the dependent variable at 1.0 reduces the
measurement error and results in a mean RJR Nabisco (IRS) gain of $341 ($323),
or a difference of 5 percent of RJR gains.

6. Regression results

6.1. Primary regressions

Table 2 presents regression coefficient estimates from examining the full sample
and the two subsamples. Inferences are similar across samples. In particular, using
conventional significance levels (5 percent, one-tailed tests), all three regressions
show compliance is decreasing in:
<table>
<thead>
<tr>
<th>Table 2</th>
<th>Summary statistics from regressions of the ratio of IRS gain to RJR gain (Ratio) on various tax return line items. IRS proceeds match RJR Nabisco proceeds within 3 percent*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pred.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>823</td>
<td>Income measures</td>
</tr>
<tr>
<td>824</td>
<td>Wages</td>
</tr>
<tr>
<td>825</td>
<td>Taxable Interest</td>
</tr>
<tr>
<td>826</td>
<td>Exempt Interest</td>
</tr>
<tr>
<td>827</td>
<td>Dividends</td>
</tr>
<tr>
<td>828</td>
<td>Pension</td>
</tr>
<tr>
<td>829</td>
<td>RJR Gain</td>
</tr>
<tr>
<td>830</td>
<td>Tax price: MTR</td>
</tr>
<tr>
<td>831</td>
<td>Socioeconomic characteristics</td>
</tr>
<tr>
<td>832</td>
<td>Homeownership</td>
</tr>
<tr>
<td>833</td>
<td>Age</td>
</tr>
<tr>
<td>834</td>
<td>Single</td>
</tr>
<tr>
<td>835</td>
<td>Male</td>
</tr>
<tr>
<td>836</td>
<td>Index</td>
</tr>
<tr>
<td>837</td>
<td>Table</td>
</tr>
<tr>
<td>838</td>
<td>Index*Table</td>
</tr>
<tr>
<td>839</td>
<td>Preparer</td>
</tr>
<tr>
<td>840</td>
<td>Aggressive opportunities measures</td>
</tr>
<tr>
<td>841</td>
<td>Sch_C</td>
</tr>
<tr>
<td>842</td>
<td>Sch_F</td>
</tr>
<tr>
<td>843</td>
<td>Rest</td>
</tr>
<tr>
<td>844</td>
<td>Other Income</td>
</tr>
<tr>
<td>845</td>
<td>Other Taxes</td>
</tr>
<tr>
<td>846</td>
<td>Other Interest Paid</td>
</tr>
<tr>
<td>847</td>
<td>Cash Contributions</td>
</tr>
<tr>
<td>848</td>
<td>Noncash Contributions</td>
</tr>
<tr>
<td>849</td>
<td>Adjusted R²</td>
</tr>
<tr>
<td>850</td>
<td></td>
</tr>
</tbody>
</table>

* See Table 1 for variable descriptions.

- several measures of income (wages, taxable interest income, pension income, and the RJR Nabisco capital gains according to the company's shareholder records);
- first-dollar marginal tax rates;
- three indicators of individual noncompliance penalty perceptions (self-employment income, rental income and other interest paid).

To elaborate, we find most of the income measures are negatively correlated with compliance regardless of sample. Our finding that absolute risk aversion is decreasing in wealth is consistent with results from Clotfelter (1983) and...
Feinstein's (1991) non-pooled tests. Our finding that compliance is decreasing in 
first-dollar marginal tax rates is consistent with Poterba (1987) and Clotfelter 
(1983). Finally, the negative coefficients on income from self-employment and 
rental activities suggest that individuals who engage in these notoriously noncom-
pliant activities are more likely to underreport capital gains. The relation between 
reporting other interest paid and capital gains noncompliance provides further 
support for cross-individual differences in penalty perception.

Several other regression coefficient estimates are significant in two of the three 
regressions. Specifically, when Ratio is truncated, we find compliance is increasing 
in homeownership, less for older taxpayers, and decreases with dividends and cash 
charitable contributions. The largest two samples also find compliance decreases 
with other taxes.

6.2. Regressions split by age

The results in Table 2 suggest that older taxpayers might be less compliant than 
younger taxpayers. To extend our analysis of the influence of age on compliance, 
we partition the sample on Age and reestimate the equation with the 181 
observations where Ratio is truncated at 1.0.16 Table 3 presents estimated 
regression coefficients for a sample of 84 older taxpayers and a sample of 97 
younger taxpayers.

We find that inferences with respect to the income variables are similar for both 
age groups and consistent with the primary regression discussed above. The 
non-income variables indicate a few differences between the groups. Specifically, 
we find first-dollar marginal tax rates only affect the compliance decisions of older 
taxpayers. Compared with younger taxpayers, we also find the old are more 
compliant if they use a paid preparer or are self-employed, but less compliant 
when they are a landlord or own their home.

6.3. Measuring compliance

Finally, we attempt to quantify the noncompliance by using a sample that 
should be free of the confounding effects of misreporting from both gifts and 
stock splits. Accordingly, we limit the sample to those observations where 
taxpayers neither overreport proceeds nor underreport basis.17 This should exclude 
taxpayers who appear to pay more taxes than they correctly owe, either by 
understating their basis or by overstating their proceeds. Consequently, attributing 
noncompliance to intentional underreporting should be more defensible for this

---

16 Results are qualitatively identical using the other two samples.
17 Note that this methodology does not impose a proceeds match, as was required in the primary 
analyses and detailed in Section 4.3.
Table 3

Summary statistics from regressions of the ratio of IRS gain to RJR gain (Ratio) on various tax return line items. IRS proceeds match RJR Nabisco proceeds within 3 percent, Ratio < 1.0, split by Age*

<table>
<thead>
<tr>
<th>Pred.</th>
<th>Age = 0 (younger)</th>
<th>Age = 1 (older)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 97)</td>
<td>(n = 84)</td>
</tr>
<tr>
<td>Income measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wages</td>
<td>−0.371</td>
<td>−7.90</td>
</tr>
<tr>
<td>Taxable Interest</td>
<td>−0.166</td>
<td>−1.91</td>
</tr>
<tr>
<td>Exempt Interest</td>
<td>0.285</td>
<td>0.54</td>
</tr>
<tr>
<td>Dividends</td>
<td>−0.376</td>
<td>−2.70</td>
</tr>
<tr>
<td>Pension</td>
<td>−0.200</td>
<td>−1.68</td>
</tr>
<tr>
<td>RJR Gain</td>
<td>−0.269</td>
<td>−8.19</td>
</tr>
<tr>
<td>Tax price: MTR</td>
<td>−0.045</td>
<td>−0.79</td>
</tr>
<tr>
<td>Socioeconomic characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homeownership</td>
<td>?</td>
<td>0.026</td>
</tr>
<tr>
<td>Single</td>
<td>?</td>
<td>−0.001</td>
</tr>
<tr>
<td>Male</td>
<td>?</td>
<td>0.001</td>
</tr>
<tr>
<td>Index</td>
<td>?</td>
<td>0.000</td>
</tr>
<tr>
<td>Table</td>
<td>?</td>
<td>−0.031</td>
</tr>
<tr>
<td>Index*Table</td>
<td>−0.001</td>
<td>−0.81</td>
</tr>
<tr>
<td>Preparer</td>
<td>?</td>
<td>−0.046</td>
</tr>
<tr>
<td>Aggressive opportunities measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sch C</td>
<td>−0.816</td>
<td>−3.85</td>
</tr>
<tr>
<td>Sch F</td>
<td>0.865</td>
<td>0.88</td>
</tr>
<tr>
<td>Rent</td>
<td>−0.061</td>
<td>−0.15</td>
</tr>
<tr>
<td>Other Income</td>
<td>0.759</td>
<td>0.40</td>
</tr>
<tr>
<td>Other Taxes</td>
<td>−0.761</td>
<td>−0.16</td>
</tr>
<tr>
<td>Other Interest Paid</td>
<td>−4.942</td>
<td>−1.46</td>
</tr>
<tr>
<td>Cash Contributions</td>
<td>−0.770</td>
<td>−2.53</td>
</tr>
<tr>
<td>Noncash Contributions</td>
<td>1.559</td>
<td>0.47</td>
</tr>
<tr>
<td>Adjusted $^2$</td>
<td></td>
<td>0.66</td>
</tr>
</tbody>
</table>

* See Table 1 for variable descriptions.

A sample of 136 observations with a mean Ratio of 0.80 than it would be for the larger samples examined above.

Table 4 presents the results of reestimating Eq. (1) for this sample. All of the original findings detailed in Table 2 hold, i.e. compliance is decreasing in wages, taxable income, pension income, the RJR gain, marginal tax rates, self-employment income, rental income, and other interest paid. We also find that, for this sample, noncompliance is greater if the taxpayer employs a tax return preparer.

Table 5 attempts to estimate the amount of noncompliance using this relatively pristine sample. We find that the mean capital gains per the RJR Nabisco shareholder records are $243 thousand (89 percent of proceeds). The mean level of noncompliance is $28 thousand (11 percent of total gains) with mean proceeds understatement of $16 thousand and mean basis overstatement of $12. At a
Table 4  
Summary statistics from regressions of the ratio of IRS gain to RJR gain (Ratio) on various tax return line items. Sample excludes observations with underreported basis or overreported proceeds deleted (n = 136)*

<table>
<thead>
<tr>
<th></th>
<th>Pred.</th>
<th>Coeff.</th>
<th>t-Stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wages</td>
<td>–</td>
<td>-0.320</td>
<td>-5.74</td>
</tr>
<tr>
<td>Taxable Interest</td>
<td>–</td>
<td>-0.430</td>
<td>-4.16</td>
</tr>
<tr>
<td>Exempt Interest</td>
<td>–</td>
<td>0.195</td>
<td>1.39</td>
</tr>
<tr>
<td>Dividends</td>
<td>–</td>
<td>-0.293</td>
<td>-1.52</td>
</tr>
<tr>
<td>Pension</td>
<td>–</td>
<td>-0.392</td>
<td>-3.22</td>
</tr>
<tr>
<td>RJR Gain</td>
<td>–</td>
<td>-0.164</td>
<td>-7.21</td>
</tr>
<tr>
<td>Tax price: MTR</td>
<td>–</td>
<td>-0.271</td>
<td>-2.49</td>
</tr>
<tr>
<td>Socioeconomic characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homeownership</td>
<td>?</td>
<td>-0.002</td>
<td>-0.08</td>
</tr>
<tr>
<td>Age</td>
<td>?</td>
<td>0.018</td>
<td>0.79</td>
</tr>
<tr>
<td>Single</td>
<td>?</td>
<td>0.015</td>
<td>0.61</td>
</tr>
<tr>
<td>Male</td>
<td>?</td>
<td>0.048</td>
<td>1.29</td>
</tr>
<tr>
<td>Index</td>
<td>?</td>
<td>0.001</td>
<td>1.21</td>
</tr>
<tr>
<td>Table</td>
<td>?</td>
<td>-0.096</td>
<td>-1.91</td>
</tr>
<tr>
<td>Index×Table</td>
<td>–</td>
<td>0.004</td>
<td>2.50</td>
</tr>
<tr>
<td>Preparer</td>
<td>?</td>
<td>-0.053</td>
<td>-2.04</td>
</tr>
<tr>
<td>Aggressive opportunities measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sch_C</td>
<td>–</td>
<td>-0.476</td>
<td>-2.44</td>
</tr>
<tr>
<td>Sch_F</td>
<td>–</td>
<td>0.937</td>
<td>0.93</td>
</tr>
<tr>
<td>Rent</td>
<td>–</td>
<td>-0.742</td>
<td>-4.85</td>
</tr>
<tr>
<td>Other Income</td>
<td>–</td>
<td>0.214</td>
<td>0.47</td>
</tr>
<tr>
<td>Other Taxes</td>
<td>–</td>
<td>-4.200</td>
<td>-1.26</td>
</tr>
<tr>
<td>Other Interest Paid</td>
<td>–</td>
<td>-6.500</td>
<td>-2.35</td>
</tr>
<tr>
<td>Cash Contributions</td>
<td>–</td>
<td>0.075</td>
<td>0.18</td>
</tr>
<tr>
<td>Noncash Contributions</td>
<td>–</td>
<td>-1.990</td>
<td>-1.69</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td></td>
<td></td>
<td>0.59</td>
</tr>
</tbody>
</table>

* See Table 1 for variable descriptions.

percent combined Federal and state tax rate and assuming no adjustments upon audit, the mean tax savings from noncompliance are $9 thousand per tax return. The 11 percent level of noncompliance detected in this study for sales of RJR Nabisco stock exceeds prior estimates of capital gains noncompliance of 7 percent in 1988 for all capital asset sales (Internal Revenue Service Research Bulletin, 1994). Because Thompson (1987) reports that stockholders are more compliant than other sellers of capital asset property, we infer that the sellers of RJR Nabisco stock during its leveraged buyout engaged in usually high levels of noncompliance for shareholders.

Readers should interpret these estimates of noncompliance with caution for at least three reasons. One, as noted above, the RJR Nabisco leveraged buyout is not
Table 5

Estimation of capital gains noncompliance by proceeds understatement, basis overstatement, and total noncompliance. Sample excludes observations with underreported basis or overreported proceeds (n = 136)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
<td>0.80</td>
<td>0.90</td>
<td>0.32</td>
</tr>
<tr>
<td>RJR Nabisco proceeds</td>
<td>271,920</td>
<td>133,101</td>
<td>399,166</td>
</tr>
<tr>
<td>IRS proceeds</td>
<td>255,929</td>
<td>120,617</td>
<td>381,981</td>
</tr>
<tr>
<td>Proceeds understatement</td>
<td>15,991</td>
<td>1843</td>
<td>51,424</td>
</tr>
<tr>
<td>RJR Nabisco basis</td>
<td>28,647</td>
<td>12,825</td>
<td>41,175</td>
</tr>
<tr>
<td>IRS basis</td>
<td>40,324</td>
<td>19,840</td>
<td>53,161</td>
</tr>
<tr>
<td>Basis overstatement</td>
<td>11,677</td>
<td>2770</td>
<td>24,655</td>
</tr>
<tr>
<td>RJR Nabisco gain</td>
<td>243,273</td>
<td>115,114</td>
<td>362,348</td>
</tr>
<tr>
<td>IRS gain</td>
<td>215,604</td>
<td>94,007</td>
<td>339,711</td>
</tr>
<tr>
<td>Total noncompliance</td>
<td>27,668</td>
<td>7554</td>
<td>58,272</td>
</tr>
<tr>
<td>as a percentage of RJR Nabisco proceeds</td>
<td>10.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total noncompliance as a percentage of RJR Nabisco gains</td>
<td>11.4%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

representative of the typical transaction that generates capital gains taxes. It was the largest (and arguably most publicized) U.S. acquisition to date.

Two, the North Carolina shareholders examined in this study are not representative of the typical RJR Nabisco shareholder. Tax bases are unusually low, and thus capital gains particularly large, providing added incentive to underreport gains. In addition, they held their stock until 1989, i.e. did not sell during the initial uncertainty surrounding the leveraged buyout. This willingness to hold shares during a turbulent period may indicate lower tax bases (as documented in Landsman and Shackelford, 1995), higher risk preferences, or less active trading among these shareholders. The impact of these factors on noncompliance is unclear.

Three, despite the similarities in the coefficient estimates from regression equation (1), the 136 taxpayers examined in this analysis are not representative of the other RJR Nabisco’s North Carolina shareholders examined in this study for at least two reasons. First, they provide unusually extensive detail in their tax returns about the RJR Nabisco transaction. To the extent incomplete and summarized data (e.g., combining all capital gains activity into one line on the return) facilitates noncompliance, the estimates based on these 136 taxpayers may understate the actual level of noncompliance. Second, our ability to match precisely these tax returns and shareholder records is consistent with these taxpayers neither receiving their stock through inter vivos gifts nor holding them in trusts or street name accounts. This suggests that unlike many North Carolina holdings, these shares
probably had not been within families for long periods. Whether this distinction affects compliance is unclear.

7. Summary

This study matches shareholder data provided by RJR Nabisco and tax return information provided by the IRS to identify the sources of cross-taxpayer variation in compliance for capital gains taxes associated with the RJR Nabisco leveraged buyout. Access to these extraordinary data permits this study to provide the most comprehensive analysis of capital gains tax compliance to date.

We conclude that:

- Compliance is decreasing in income, consistent with decreasing absolute risk aversion.
- Compliance is decreasing in the size of the capital gain, as determined using RJR Nabisco shareholder records.
- Compliance among older taxpayers is decreasing in first-dollar marginal tax rates.
- Compliance is increasing in the perceived penalty on detection.
- On average, taxpayers did not report 11 percent of total gains.

The findings in this paper advance our understanding of capital gains tax compliance and should aid policymakers, and taxing authorities in forecasting compliance as investors unwind appreciated positions created by the bull market of the 1990s.

In closing, let us repeat that though the uniqueness of this data base enables us to analyze capital gains tax compliance with unusual power, it also limits the generalizability of the results. The sample is composed of shareholders of RJR Nabisco who sold their shares during its 1989 leveraged buyout and provided sufficient information on their tax returns for us to match their tax returns to their shareholder records. It is impossible for us to assess whether the compliance levels of these shareholders represent typical equity investors.

Acknowledgements

This paper has benefited from comments by two anonymous referees and Jim Poterba (editor), financial support by the Center for Finance and Accounting Research at UNC-Chapel Hill, the Dalton L. McMichael, Sr. Professorship (Landsman), and the Jefferson-Pilot Faculty Development Fund (Shackelford), discussions with Ramji Balakrishnan and Charles Christian, data support from RJR
Nabisco and the Internal Revenue Service, and research assistance by Susan Eldridge.

References


