NBER WORKING PAPER SERIES

TAXPAYER SEARCH FOR INFORMATION:
IMPLICATIONS FOR RATIONAL ATTENTION

Jeffrey Hoopes
Daniel Reck
Joel Slemrod

Working Paper 19482
http://www.nber.org/papers/w19482

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2013

We wish to thank John Guyton and Ron Hodge of the Research, Analysis, and Statistics Division of the Internal Revenue Service, for helping us use the IRS aggregate administrative data. The views expressed here are those of the authors alone, and do not reflect the views of the Internal Revenue Service or the National Bureau of Economic Research. We are grateful for comments on an earlier draft received from Hal Varian, Raj Chetty, Alex Gelber, and participants at the University of Michigan Hosmer - Hall Luncheon series, the US Treasury Department Office of Tax Analysis, the University of Michigan Public Economic Summer Talks series, the University of Uppsala, and at the 2013 Public Economics United Kingdom Conference.

Taxpayer Search for Information: Implications for Rational Attention
Jeffrey Hoopes, Daniel Reck, and Joel Slemrod
NBER Working Paper No. 19482
September 2013
JEL No. D80,D83,H24,H31

ABSTRACT

We examine novel data on searches for capital-gains-tax-related information to determine when and how taxpayers acquire information. We find strong seasonal increases in information search around tax filing deadlines, suggesting that taxpayers seek information to comply with tax laws. Positive correlations between stock market activity and information search and year-end spikes in information search on capital losses suggest that taxpayers seek information for tax planning purposes. Policy changes and news events cause noteworthy information search. Overall, these data suggest that taxpayers are not always fully informed, but that rational attention and exogenous shocks to tax salience drive taxpayer information search.

Jeffrey Hoopes
Ohio State University
448 Fisher Hall
2100 Neil Avenue
Columbus, OH 43210
hoopes@fisher.osu.edu

Daniel Reck
University of Michigan
Economics Department
238 Lorch Hall, 611 Tappan Street
Ann Arbor, MI 48109-1220
dreck@umich.edu

Joel Slemrod
University of Michigan Business School
701 Tappan Street
Room R5396
Ann Arbor, MI 48109-1234
and NBER
jslemrod@umich.edu

1. Introduction

Taxes, especially income taxes, can be complex and confusing. Despite a general awareness of this fact, the consequences of complexity and misunderstanding are not well-understood. Survey evidence suggests that many taxpayers do not understand basic tax concepts (Blendon et al, 2003), and the compliance cost of taxes, including learning enough about them to comply, is large (Slemrod, 1995; 2004). Given confusion surrounding tax incentives and tax law, many taxpayers might simply ignore or misperceive the incentives built into the tax code when they make decisions with tax consequences. Alternatively, taxpayers might collect information on the taxes that matter to them, and use this information to make tax-efficient decisions. Finally, taxpayers might learn just enough about tax policies to fill out their tax return and perhaps avoid an audit.

In this paper, we investigate these alternative views of taxpayer information search. We reject that taxpayers are fully informed about the tax system, that they act in complete ignorance of the tax system, and that they gather information only for tax compliance purposes. The evidence we present suggests instead that taxpayers employ rational attention to tax policies, in line with theories proposed by Sims (2003) and Reis (2006). In addition, we observe that exogenous shocks to tax salience from news events can substantially increase information search.

Modern technology has greatly expanded the accessibility of information. Any person with access to the Internet may, in a few minutes, learn at least something about the most obscure details of the tax code. Taxpayers undoubtedly do use these resources to seek information: people Google "tax" more often than they Google the names of public figures,[1] the IRS website has received on average 4.6 million visits *per day* since 2004, and the IRS call line has received on average 125 thousand calls per day since 1999. How tax knowledge matters hinges on 1) how

---

[1] For example, Google searches for "tax" were more common than searches for "Barack Obama," and were more common than "Obama" in all but the 2008 and 2012 election season.

and when people seek out tax-related information, and 2) whether they change their behavior once they acquire it. In this paper, we address primarily the first of these questions, and provide some preliminary analysis of the second.

Our data consist of aggregate high-frequency time series on calls to the IRS toll-free phone number, aggregate visits to the IRS website, measures of Google searches on tax-related terms, and views of tax-related web pages on Wikipedia. We also use supplementary data on searches conducted on the IRS website and referring pages to the IRS website. From these sources we collect data on information seeking regarding one specific tax-related topic: capital gains taxes. We select this topic because it is a perennially controversial policy issue, because data on the relevant taxed behavior, sales of capital assets, are available on a high-frequency basis, and because the American Taxpayer Relief Act of 2013 (ATRA) enacted a change in capital gains tax rates during our sample period.

We use data on information search to shed light on competing theories of information and attention, and in particular different assumptions on the role of information in decision-making. We study information-seeking around five different types of events: 1) time notches, 2) macroeconomic changes, 3) policy changes or the mention of potential policy changes, 4) filing deadlines or approaching filing deadlines, and 5) tax-related news events.

Information search around the first three dimensions should be tied to the taxpayer's ability to make fully informed decisions affected by tax incentives. First, a particular date is often relevant for the incursion of tax liability or the tax efficiency of behavior. We call these time notches to indicate that tax liability can change abruptly, and non-discretely, at certain dates, usually at year-ends. For example, the last date to affect one's capital gains tax liability in a given year is December 31st. As the deadline approaches individuals may research the benefits from realizing a capital gain or loss in the current year as opposed to a future year. Second, macroeconomic changes may be associated with information search due to perceived changes in the importance of understanding the tax consequences of a particular behavior. For example, asset owners may seek information about capital gains taxation as the stock market falls to understand the tax implications of their losses. Third, actual or potential changes in tax policy or tax enforcement may cause individuals to learn about the tax system, either to plan for the future or to make an informed voting choice. Taxpayers might also seek information on the tax system when completing their tax return. As such, the fourth dimension we study is filing deadlines and the onset of these deadlines, when individuals may learn about taxes due to approaching deadlines for filing tax returns. For the taxes we consider, the deadline affecting most taxpayers is in mid-April.

News events sometimes spark public interest in tax policy, and in this case individuals may search for tax information in order to develop an informed opinion on current events or out of curiosity. The fifth dimension we study consists of news stories such as the release of a public person's tax information. We can view these events as exogenous shocks to the salience of a particular aspect of tax policy. They have no direct bearing on an individual's tax burden, nor do they directly affect tax incentives. The search may, however, provide information that incidentally informs the searcher about her own decisions.

We observe strong seasonality in the search for information on capital gains taxes through all channels. Taxpayer information search increases substantially during the period commonly called "tax season," which runs from mid-January to mid-April of each year. An even more pronounced spike in information search occurs very close to the filing deadline in mid-April. We also document the impact of several discrete events on taxpayer information search regarding capital gains taxes, through Google and Wikipedia. Presidential debates in which candidates discuss their proposals for capital gains taxes, the passage on January 2, 2013 of ATRA, the release of then presidential candidate Mitt Romney's 2010 tax return in January of 2012, presidential elections, and policy changes all generate large and significant increases in taxpayer information search. In every case, these events cause a spike in taxpayer information search that fades within three to four days.

Next, we find that macroeconomic changes affect information search on capital gains taxes through Google and Wikipedia. We observe significantly elevated information search on days with large trading volume in the stock market. We also explore a measure of information search for personal investment advice generally, using Google searches. Daily searches for stock advice predict searches for capital gains tax on the same day, the previous day, and, in some specifications, one day in the future. This is the first evidence we find, and in the literature, that some taxpayers investigate the tax consequences of an action while contemplating the action itself.

These patterns indicate that information search is consistent with a model in which individuals search when different events make understanding tax policy more important, i.e. when they increase the return to information search. However, these results are also consistent with individuals making choices and *then* researching the tax implications of those choices for their wealth. In an effort to provide evidence that more strongly suggests a causal relationship, whereby events cause individuals to search for information in order to make more informed decisions, we examine information search on capital losses. We document substantial elevation in information search related to capital *losses* at year-ends, especially in 2008 (when capital losses from the stock market crash would have made taxpayers' more likely to be rebalancing their portfolios in order to take tax-efficient advantage of capital losses). Observing increases in information search just prior to the time notch is consistent with taxpayers seeking information on the tax system and using it to improve decision-making.

This paper provides the first-ever attempt at understanding how, and why, taxpayers search for information about government policy, specifically tax policy. We establish that taxpayers seek information in order to both comply with their tax obligation and to respond behaviorally to tax incentives. Taking taxpayer learning seriously has important implications for understanding the impact of tax policy and tax policy changes as well as for a full understanding of the nascent idea of tax salience.

2 Literature Review

2.1 Public Understanding of Taxes

Political science research is fairly persuasive that voters know very little about the details of government. Delli Carpini and Keeter (1996), in a comprehensive survey of the political knowledge of voters covering several decades and hundreds of surveys, show that the majority of voters are ignorant of many key aspects of the U.S. political system. Surveys suggest that people are also largely ignorant of the tax system (Blendon et al (2003) summarize the results).

One specific setting where researchers have investigated understanding of the tax system is taxpayers' perceptions of their average and marginal tax rate. Brown (1968) and Fujii and Hawley (1988) find that individuals' self-reported marginal tax rate often differs from the true rate that can be estimated from their demographic characteristics. Research by de Bartoleme (1995) shows that, in a lab experiment, MBA students often confuse the average tax rate with the marginal tax rate when making investments in a taxable versus non-taxable project. Sheffrin (1994) and Liebman and Zeckhauser (2004) also find evidence that taxpayers systematically misunderstand income tax schedules.

Recently, the public economics literature has settled on the term "salience" to capture the extent to which tax aspects of the environment are noticed, and acted upon, by those affected. A key paper in the modern literature is Chetty, Looney, and Kroft (2009), who in a field experiment find that posting at the store shelf retail-sales-tax-inclusive prices, rather than adding the tax liability at the point of sale, reduces purchases. Their findings comes from an experiment at a large grocery retailer in California, at which prices inclusive of the 7.375 percent state sales tax were posted alongside the original pre-tax price over a three-week period for three product categories (cosmetics, hair care accessories, and deodorants). They estimate that the "tax treatment" reduced demand by 8 percent; given demand elasticities of 1 to 1.5 for the affected products, they conclude that most consumers do not take into account the sales tax revealed at the cash register.

2.2 Macroeconomics and finance

Many macroeconomists have considered how best to account for the inertia in observed economic behavior and to what extent imperfect information can account for it. For example, the models in Sims (2003) are motivated by the idea that information that is freely available to an individual may not be used, because of the individual's limited information processing capacity. Alternatively, in Reis (2006) consumers rationally choose to only sporadically update their information and re-compute their optimal plans, while in between updating dates they remain inattentive. Both models imply that news disperses slowly throughout the population, so events have a gradual and delayed effect on behavior.

A recent literature in financial economics has taken advantage of newly available data and focused on the demand by investors for information. Da, Engelberg, and Gao (2011) propose the Google Search Volume Index (GSVI) as a direct measure of investor attention. Drake, Roulstone and Thornock (2012) use the GSVI for public company ticker symbols to examine the timing and magnitude of Internet search around earnings announcements and the factors that influence Internet search. Among their results of interest is that investors extend more effort when the potential returns to search are higher.

2.3 Testable implications of theories

Although our analysis focuses on taxpayers' search for information about income taxes, it has more general implications regarding the information economic agents possess when making decisions (in our case, often very economically large decisions). Do individuals possess full information for important economic decisions? If not, is attention to important information rational? If attention is rational, how should we model information acquisition? We will shed some light on the answers to these questions using patterns in the timing of information search. To the best of our knowledge, all previous economic studies on attention have assumed that attention itself is unobservable, but we take the approach of directly observing attention and testing whether it conforms to the assumptions embedded in the theory. Table 1 summarizes the predictions of the theories our analysis illuminates, as described in this section.

The null hypothesis in our analysis is that taxpayer information search is unresponsive to political or economic events. This hypothesis would be implied by the assumption of full information, wherein an individual always has all data necessary to make an optimal choice. Although frequently criticized and relaxed in the economics literature, this assumption is still common in public finance for modeling the response to tax incentives and the welfare cost of taxation. It stands to reason that fully informed individuals should not search for information. Why spend time researching something that is already completely understood?

In rational expectations models like the one proposed by Muth (1961), the individual faces uncertainty about the future, but knows everything about the present and costlessly and instantaneously absorbs all new information. Rational expectations might allow searches for tax information to respond instantaneously to unanticipated policy changes or shocks to the probability of a policy change, as these require that the individual absorb new information about policy. However, information search should not respond under rational expectations to economic or political events that change the importance of understanding tax policy, but not tax policy itself. Information search should also not respond to the implementation of policy changes that were not surprises. Individuals should already possess all publicly available information on tax policy.

An attractive alternative framework is rational attention, under which individuals have limited capacity for processing new information and they allocate that resource optimally. Alternative assumptions for how limited attention may be rationally allocated are proposed by Sims (2003) and Reis (2006). In either model, information search responds positively to events that increase the utility gain to understanding the incentives at play in making a decision (such as a market downturn or approaching time notch) and to events which introduce new information about incentives (such as policy changes).

One difference between these two models concerns the precise timing of information acquisition. In Sims' (2003) approach, agents constantly update their beliefs based on their attention to different sources of information, so we should expect information search to instantly react to events that change the importance of attending to a given bit of information. In Reis' (2006) approach, updating beliefs is costly and upon updating the individual simultaneously plans future consumption and decides when to update again. In between updating dates, the individual is inattentive to new information. Events like policy changes should affect information search in the Reis (2006) model, but the searching should be spread out over a longer time period, as not all individuals constantly update beliefs. We can therefore use the timing of information search around unanticipated economic and political events to test the difference between the mechanisms for rational attention employed by the two models. A near instantaneous response indicates that the data favor the Sims (2003) mechanism over the Reis (2006) one.

Attention is also the subject of a large literature in psychology. One useful distinction in this literature is between "exogenous," or "stimulus-driven," attention and "endogenous," or "goal-directed," attention (Theeuwes, 1994; Connor, Egeth, and Yantis, 2004). Endogenous attention is the same as rational attention: the individual voluntarily directs attention to meet a goal. Exogenous attention is driven, rather, by an external stimulus, such as increased salience. In this case, the individual might search for information regarding anything she sees in the news which she does not understand, regardless of her gain from the search. If attention to economic information is exogenously driven at times, information search should respond to news events that mention policy—because these are usually accompanied by news coverage that makes the policy more salient.

3 Capital Gains Taxation and Behavior

Capital gains generated from the sale of capital assets have received preferential treatment relative to labor income since 1921 in the U.S. Income from the sale of capital assets is recognized in the year of sale, and the taxable income is equal to the sale price of the asset less its tax basis (the historical price plus any acquisition costs and improvements to the asset, less any accumulated depreciation). Capital gains and losses are divided into two categories, short-term and long-term. Long-term (short-term) capital gains and losses arise from the sale of a capital asset that has been held for more than (less than) one year. Long-term capital gains receive a favorable tax treatment and are currently taxed at a maximum of 20%. Short-term gains are taxed as ordinary income. Gains and losses from short and long term assets are netted together in well-defined ways to determine a taxpayer's ultimate tax liability.

A rational investor should maximize after-tax utility, which would entail taking capital gains taxes into account in deciding whether to buy, sell, or hold a capital asset. First consider the purchase decision in isolation from the rest of an investor's portfolio. The taxation of the capital gains tax structure reduces the expected after-tax rate of return to a capital investment, where the reduction depends on the expected appreciation, the expected holding period, and the likelihood that an asset with appreciation can be held until death. Because higher capital gains taxes reduce the attractiveness of assets expected to appreciate in value, one may expect that the level of asset prices would react negatively to unexpected news about tax increases (described as the capitalization effect).[2]

The decision of if and when to sell a capital asset should also be affected by the tax system. Certain provisions in the tax code (for example, a lower tax rate for long term capital gains), may encourage taxpayers to postpone an asset sale in order to obtain the favorable tax treatment. Likewise, the annual nature of tax compliance may also create annual rebalancing of portfolios to achieve a favorable mix of capital losses and gains (such that the losses almost exactly cancel the gains), and may encourage asset sales.

Because capital gains taxes are triggered by asset sales that happen in year t, there is a planning deadline for capital asset sales at the end of the year. Therefore, information gathering in order to achieve the best after-tax outcome with regards to capital asset sales will have to happen by December 31 of year t. When capital gains tax rates increase between year t and year t+1, December 31 is also the last date to realize capital gains at the year-t tax rate. As such, information gathering in order to shift capital income through time in response to a rate change will also need to happen by December 31st of the end of the year.

Previous empirical evidence is largely consistent with taxpayers altering both the nature and the timing of transactions in order to achieve the greatest possible after-tax return on their capital investment. For example, as capital losses must generally be offset with capital gains, large downward movements in the stock market often leave investors with large unrealized capital losses that must be carried forward until years when these investors realize capital gains. This policy leads to the well-known strategy for minimizing capital gains tax liability of "loss harvesting," selling capital assets with built-in losses to offset the gains of assets you have sold in the year. For example, Ivković, Poterba and Weisbenner (2005)

---

[2] Of course, the effect of acquiring information about capital gains taxes depends on how it changes prior beliefs. One can imagine a potential investor being pleasantly surprised to learn about the preferential lower tax rate and step-up basis at death, or being discouraged upon learning that any tax at all applies upon the sale of appreciated assets.

find evidence of tax-loss selling at the end of the year. Poterba and Weisbenner (2001) find that this is especially prevalent in years when changes in tax policy provided additional incentive to harvest losses at year-end. While tax-loss selling occurs in the corporate equity market, there is also evidence of such activity in municipal bond closed-end funds (Starks, Yong, and Zheng, 2006) and the market for long-term government and corporate bonds (Chang and Pinegar, 1986).

4 Data

Tax information is made available by the IRS, by accounting firms and other organizations in the business of providing information generally, such as Wikipedia. The IRS-provided information can be accessed through the Internet, via printed information booklets, and through toll-free numbers. Person-specific information can also be obtained by perusing one's paycheck, one's Form W-2, or one's prior tax returns. Both information from the IRS website and information from non-government organizations may be found quickly using Google searches. In all cases, the marginal cost of public information is only the time spent acquiring it, provided one has an Internet or phone connection. The cost of tax information decreased with the advent of the Internet, and search engines in particular. Long-run trends in information search may therefore vary due to both supply and demand for information. By focusing on high-frequency variation in information search, we isolate variation in the demand for information, since the supply changes little from day to day.

We examine measures of information search through as many channels as is feasible. We first study data on tax-related information search from Google, accessed via Google Trends. One can use Google searches to find information from a variety of sources, including Wikipedia and the IRS. Using query data on Google searches, Google Trends provides a measure of the "propensity to search" for a given search query or set of queries. More specifically, an observation in the Google Trends data will be, for a given day and geographical region, the number of Google searches for the specified search terms divided by the total number of Google searches on any topic in the time and place.[3] For our purposes, the search terms will be a broad set of capital-gains-tax-related search terms, and the geographic region is the United States. After calculating the propensity to search, Google scales it from 0 to 100, where the number 100 represents the day with the highest search volume for this set of search terms in the entire sample period. Our sample period consists of January 1, 2004 (when data are available) through March 30, 2013.

Entering a given term into Google will frequently yield a Wikipedia page related to that term as one of the first suggested sites to visit. Our second source information search data is page view data from Wikipedia, a free online encyclopedia edited by Internet users. We obtain hourly page views to the English language Wikipedia site, "Capital Gains Taxation in the United States." We add these over the 24 hours in a day to obtain daily data for January 1, 2008 through March 30, 2013. One disadvantage to the Wikipedia data is that it is available for a substantially shorter time period than is the Google data. On the upside, we are able to obtain the raw number of views of the webpage, which simplifies the interpretation of the size of changes in information search.[4]

Our third source of data regards aggregate calls made to the IRS's toll free phone number, where taxpayers can call and speak to a representative from the IRS or listen to automated messages. In the course of a phone call with the IRS, some taxpayers will listen to prerecorded messages about various "tax topics." We analyze daily phone calls that access the IRS tax topic 409, "Capital Gains and Losses."[5] Taxpayers may reach the tax topics function in a number of ways, including calling the tax-topics-specific phone number, or being referred to the

[3] For more detail about Google Trends, see our web appendix and http://support.google.com/trends/?hl=en.
[4] We have replicated the results in Section 6 and Section 7 using a measure of Wikipedia usage scaled in the same way as the Google data, using page views of the Wikipedia home page as a metric of overall Wikipedia usage. This replication suggests strongly that scaling is not confounding our results.
[5] Other example of tax topics are tax topic 304, "Extensions of Time to File Your Tax Return" or tax topic 151, "Your Appeal Rights."

topic after a conversation with a representative or an interaction with the automated system. We use data from February 1, 2002 to March 31, 2012.[6] We also obtain the total number of calls to the call line for individuals, which is not specific to capital gains.

The last data set we analyze also comes from the IRS, and consists of the aggregate daily number of visits[7] to the IRS's webpage. The IRS maintains a website, IRS.gov, which hosts a vast amount of tax information. We obtain, directly from the IRS, the total number of page views and visits, by day, to any site hosted by IRS.gov. We are able to obtain this data for the entire time series for which the IRS has maintained the data, February 1, 2002 to March 31, 2012. Analysis of IRS.gov traffic carries the benefit of analyzing taxpayer information search about a very broad set of tax information. This measure of information search is not limited to information search about capital gains taxes specifically.

5. Estimation Procedure

Let I_{it} denote information search on date t from a source of information i, such as Google, Wikipedia, or the IRS call line. We wish to estimate the effect of several different events on information search through source i. The events we study may be either non-recurring, as in the case of a Presidential debate that mentions capital gains taxes, or may occur annually, as in the case of tax filing season. Non-recurring events are indexed by the subscript k, and annual events are denoted according to the day of the year on which they occur, denoted DoY.

As we discuss in more detail later, the data on inquiries display marked seasonality at the yearly and weekly levels. In order to evaluate whether a particular (non-seasonal) event increased information search on date t, we must properly

[6] For a limited number of hours, Wikipedia data are unavailable. We scale the observation up by 24 divided by the number of hours present in a day when only some hours are missing (15 days), and we use the previous day's value when the entire day is missing (128 days).

[7] In contrast to page views, visits are a web analytic that count the number of page views from unique Internet Protocol (IP) addresses in a given time frame.

specify the counterfactual level of search, conditional on the day of the week and day of the year of date *t*. For example, suppose we want to know the impact of an event occurring on March 1st, 2010. Suppose further, as we show later, that searches are elevated in early March, compared to other times of year (because March occurs during tax filing season). If we simply regressed information search on a dummy variable for March 1st, 2010, we would then tend to over-estimate the impact of this event if we did not control for yearly seasonality. We control for the day of the week for the same reason: all of our measures of inquiries are higher on weekdays than on weekends, so a consistent estimate of the impact of an event should control for whether that event occurred, for example, on a Monday or a Saturday.

Unlike for much analysis of time-series data, yearly seasonality is of intrinsic interest for our research questions. In particular, increases in information-seeking during tax season will affect the form of the yearly seasonality. Information searches occurring at the end of the year can be thought of as response to a time notch. To address this concern, we employ a method that both estimates the seasonal patterns and allows us to perform classical statistical tests for whether information search is significantly higher on a particular day of the year than the average level of information search. Specifically, for each information search series *i* we estimate the function:

$$I_{it} = \sum_k \left[\beta_{k0}F_{kt} + \beta_{k1}F_{k(t-1)} + \cdots + \beta_{k4}F_{k(t-4)}\right] + x_t^{'}\gamma + f_i(DoY_t) + \delta_{DoW,i} + u_{it} \quad (1)$$

The term inside square brackets is a set of dummy variables: $F_{k(t-j)}$ equals 1 if event *k* occurred on date *(t-j)* and zero otherwise. The four-day event window was selected because when large spikes in information occur in the data, search levels return to baseline within four days. We also include a vector of continuous, time-varying linear covariates x_t'—such as trading volume on the stock market—along with a general non-linear function in day-of-the-year $f_i(DoY_t)$, and a day-of-the-week fixed effect, $\delta_{DoW,i}$. We use data on $f_i(DoY_{t-1})$ and $f_i(DoY_{t+1})$ (the day before and the day after day t) to estimate $f_i(DoY_t)$, under the assumption that $f_i(DoY_t)$ should not change *too* sharply from day to day.[8]

We proceed by employing a kernel-weighted local linear regression. The strength of this estimator is that 1) it estimates a smooth trend in day-of-the-year, which allows us to control for day-of-the-year when estimating the effect of events, and 2) it increases the precision of the estimates of $f_i(DoY_t)$ relative to fixed effects. A potential downside to the estimator is that the assumption that $f_i(DoY_t)$ does not on average change *too* sharply from day to day may be wrong (especially around April 15th). The strength of this assumption is governed by the bandwidth of the estimator, and also by our choice of kernel density function. We discuss these choices below. Our preferred choices for kernel density function (Gaussian) and bandwidth (four days) produce estimates of seasonal trends that match the pattern implied by fixed-effects estimates. To avoid conflagration of seasonal effects with large outliers due to the events we document, we estimate all the components of equation (1) in a single regression. The online appendix to this paper outlines the estimation procedure in more detail. For a recurring annual event such as a filing deadline, we can examine the function $f_i(DoY_t)$ on days-of-year corresponding to the annual event to understand the effect of the event on information search. The comparison that seems most natural is to test the hypothesis that, on a particular day of the year, DoY', inquiries are higher than their unconditional average, i.e. that $f_i(DoY') > E(f_i(DoY))$.[9]

Our estimate of the function $f_i(DoY)$ will inform us about the importance of tax season for information search, but it may not capture all variation due to compliance deadlines.[10] For example, the mid-April filing deadline does not occur

[8] Alternatively, we could estimate $f_i(DoY_t)$ using day-of-the-year fixed effects. However, with only 5 to 9 years of data for each type of inquiry, these fixed effects would be imprecisely estimated.

[9] This is the same hypothesis test we conduct by including day-of-the-week fixed effects, and inquiries on the "left-out" DoY was a date on which searches were on average equal to the average over all days.

[10] For Google data, the interpretation of the estimates is somewhat complicated by the fact that the dependent variable is a measure of propensities to search. If there were yearly seasonality in overall Google searches, this would affect our estimate

on April 15 of a given year if that date falls on a weekend or Emancipation Day. Instead, it can occur as late as April 18 in some years. We account for variation in compliance dates by year by including dummy variables for compliance events in addition to the locally linear function. Our estimate of the function $f_i(DoY)$ will estimate the average effect on April 15, but we add an additional dummy variable for the precise date of the filing deadline to capture variation that occurs specifically on the deadline each year. Doing so does not significantly change our picture of the importance of tax season for inquiries, but it does highlight the sharp spike that occurs exactly on the mid-April deadline each year. The inclusion of this dummy variable also helps mitigate the smoothness assumption in estimating $f_i(DoY)$, effectively imposing a bandwidth of zero on the date of the filing deadline. To further reflect that the smoothness assumptions should not apply when there are sharp changes on particular days of the year, we also include dummy variables for the following holidays: New Year's Day, President's Day, Martin Luther King, Jr. Day, Memorial Day, Independence Day, Labor Day, Thanksgiving, and Christmas.

6 Exploring the Raw Data

Figure 1 plots the evolution of three measures of information search on capital gains taxes over time. Calls to the IRS hotline inquiring about capital gains taxes occur almost exclusively during tax season, from mid-January to mid-April. The absolute volume of calls occurring during a tax season diminishes considerably throughout the time period, from 2002 to 2012. Wikipedia page views, in contrast, increase over time. These patterns are consistent with online information largely having supplanted telephone information as Internet access increased markedly over the sample period. The Google measure displays a slight downward trend.[11]

---

of $f_i(DoY)$ by decreasing it on days of the year when Google usage was highest. The fact that the pattern estimated by Google data and other measures of information search are similar largely alleviates this concern. The inclusion of day-of-the-week fixed effects also makes our results robust to weekly seasonality in Google searches.

[11] Recall that this does not mean that Google searches of capital gains taxes decreased over time, only that the share of Google searches that concerned capital gains taxes decreased over time.

Visually apparent in each time series is a strong pattern of yearly seasonality. This is most pronounced in the IRS call log series, where the yearly seasonality drives virtually all of the variation in the series over time, but the pattern is also present in Google searches and Wikipedia page views. Zooming in on a narrower time frame also reveals strong weekly seasonality in each time series.

To focus on high-frequency variation, we detrend the data using a Hodrick-Prescott filter.[12] To allow for visual comparison of the three measures of information search, we normalize each variable by dividing by its standard deviation.[13] Figure 2 plots the estimated yearly seasonality in the standardized data.[14] For each measure of information search, we observe a sizable and significant increase in search behavior during tax filing season, and an even stronger spike in the immediate run-up to the April 15 filing deadline. Right after April 15, information search drops off sharply. The estimated seasonality in Figure 2 also matches patterns in the timing of filing documented by Slemrod et al (1997). Clearly, the desire to comply with the tax law as one fills out a tax return leads taxpayers to search for information.

In the next set of figures, we focus more narrowly on variation in the standardized, detrended, and seasonally-adjusted[15] series around three dates where we observe the largest spikes in information search. Figure 3 focuses on the spike in October and early November 2008. There are two obvious candidates for elevated information search in this time period: the presidential election of 2008, and the stock market crash. Because capital gains taxes were an issue on which candidates John McCain and Barack Obama differed substantially, the election

[12] Whenever we use a Hodrick-Prescott filter in this paper, we use a smoothing parameter of 10^7. This value was selected by trial and error, with the goal that the long-run trend capture long-run movements in the series but not variation due to yearly seasonality.

[13] Because the volatility of the IRS call log series varies significantly over time, we standardize it by dividing by its standard deviation *by year*.

[14] This estimate of seasonal effects comes from the regression in Section 7, and controls for event-driven outliers.

[15] Seasonal adjustment consists of subtracting out estimated day-of-the-week effects, holiday effects, and the day-of-the-year effect.

could influence search behavior due to either the desire to make an informed choice about asset purchases or sales based on expected future tax policy, or due to the desire to understand the consequences for future tax policy of a President McCain or Obama. The market crash of 2008 resulted in large capital losses for many investors and, due to the extreme volatility of the market, potentially also large short-run gains. Our conclusion is that both the election and the crash played a role, although it is difficult to disentangle the effect of the two events. Our conclusion is based on the fact that the largest swings in the stock market in this period (marked by red and black vertical lines in figure 3) are associated with spikes in search behavior, and that information search surged around the date of the presidential election. We present more evidence on how the crash might have affected information search by focusing more narrowly on capital losses in section 8, and Figure 6 provides further evidence on presidential elections.

Panel A of figure 4 plots the same series in January and February of 2012. We attribute the observed surge in information search to the release of presidential candidate Mitt Romney's 2011 tax return on January 24. The release generated substantial news coverage, in part due to his low effective tax rate. As most news articles on the subject noted, much of Romney's income came from the realization of long-run capital gains taxes, taxed at 15 percent. Another possible explanation for the spike in search behavior on this date is the State of the Union Address, which also occurred on January 24, 2012. In his speech, President Obama advocated taxing the wealthy at higher tax rates (supporting the "Buffett Rule"), but he mentioned neither capital gains taxes specifically nor investment taxes more generally.[16] It is, however, possible that Obama's rhetoric on taxing the rich led the public to pay greater attention to the news about Mitt Romney's tax return, and

[16] The "Buffett Rule" is a tax plan proposed by President Obama, wherein individuals making over $1.0 million in taxable income would be subject to a minimum average tax rate of 30%.

thereby amplified its effect on capital-gains-tax-related information search.[17]

Panel B of figure 4 plots the data from detrended and seasonally adjusted data at the end of 2012 and the beginning of 2013. A presidential debate between Barack Obama and Mitt Romney appears to have sparked considerable information search. As in 2008, we also see elevated search immediately following the presidential election, perhaps as voters and investors researched what might happen to capital gains taxes in the aftermath of President Obama's re-election. Finally, the largest outlier in information search in the 10-year period covered by our data—an increase of just over 4 standard deviations in Wikipedia page views and over 6 standard deviations of Google searches—occurred on January 2, 2013. We attribute this to the passage of the American Taxpayer Relief Act on that date. This bill temporarily resolved what was commonly called the "fiscal cliff" debate, and increased the top marginal tax rate on long-term capital gains from 15 percent to 20 percent. Our evidence strongly suggests, therefore, that individuals search for information both in response to policy changes and in response to potential policy changes signaled by political events.

7 Regression Analysis of the Impact of Events

We start with the detrended and seasonally *unadjusted* time series. The regression procedure we use, outlined in Section 5 and in the appendix, explicitly controls for variation due to weekly and yearly seasonality. The notes to Table 2 describe the events we study, which are also discussed in the previous section.[18]

Table 2 describes the estimated impact of events on information search through

---

[17] Further insight into this issue can be gained by examining Wikipedia intra-day search activity. Search activity started rising dramatically mid-day on January 24th, 2012, consistent with Romney's tax return being at least partially responsible for this increase (as Obama's speech was not delivered until that night).

[18] There is modestly elevated information search via Google, Wikipedia or both, on a few dates that we do not include in the analysis. These dates include January 18, 2004 (a State of the Union address delivered by George W. Bush advocating the extension of capital gains tax cuts), November 2, 2004 (re-election of George W. Bush), May 16, 2006 (the extension of the 2003 capital gains tax cuts), January 27, 2010 (a State of the Union address by Barack Obama advocating a cut in capital gains taxes for some taxpayers), and March 23, 2010 (passage of the Affordable Care Act, which included a "net investment income tax"). In each case, the response of searches is qualitatively similar—a spike in searches that fades in three to four days—but notably smaller than the events we do include in the formal analysis.

Google, Wikipedia, and the IRS web page. To examine statistical significance, we perform F-tests on the β's from equation 1, which generates a p-value that corresponds to the probability that changes in information search of the magnitude we observe would have occurred at random during the event window we specify. The results suggest that the events we study each have a large and significant impact on information search through Google and Wikipedia, but not through the IRS call line.[19] When we estimate the overall impact of the event rather than examining a single-day impact, the release of Mitt Romney's tax return surpasses the passage of ATRA as the one-time event that generated the most taxpayer information search, through both Google and Wikipedia. This occurs because the effect of ATRA passage declined more rapidly, fading in two days instead of four. Notably, ATRA passage was anticipated in the days leading up to January 2. The bill passed Congress on January 1st, and the anticipated increase in the top capital gains rate, likely generated income shifting from 2013 to 2012, which may have also caused information search. From Figure 4, panel B, we can see that elevated information search occurred on the two days prior to ATRA passage, December 31 and January 1st. Adding these dates to the event window increases the estimated impact of ATRA by approximately 5 standard deviations for Google and Wikipedia, but changes little else.

8 Relating Tax Information Search to Stock Market Activity

An advantage of focusing on capital gains taxation is the availability of high-frequency data on sales of stock. Relating these data to data on information searches holds the promise of better understanding the causal connections between information search and capital-asset-related behavior. After all, we are not only interested in what causes people to search for information, but also in to what extent

[19] Two events are estimated to have an effect on calls to the IRS. However, seasonal variation in the volatility of the calls measure means that the statistical significance procedure is biased for dates occurring during tax season. As such, we are cautious about interpreting this episode.

the acquisition of information affects behavior, in this case behavior related to the sale, purchase, and holding of capital assets.

8.1 The Lead-Lag Relationship to Volume, Volatility, and Market Return

Our first strategy is to examine the lead-lag relationship between measures of behavior and information searches. If searches lead behavior, then we have reason to pursue the idea that the information obtained affected subsequent decisions. We investigate two data series, both of which represent general stock market activity.

The market-related measure we use is trading volume. We obtain the dollar value of shares traded from all publicly listed firms from the Center for Research in Security Prices (CRSP variable VOL), which we use as a measure of broad market activity.[20] We then detrend the measure using a Hodrick-Prescott filter, and include the log of daily trading volume in the regression described in equation (1). For our baseline regression, we include five days of leads and lags of log trading volume. These regressions control for the events we study and for weekly and yearly seasonality. Columns (1) through (3) of Table 3 report the coefficients on standardized data. See Appendix Figure A.1 for a graphical depiction of the lead-lag estimates.

When trading volume is high, individuals seek more information about capital gains taxes. The standard deviation of log trading volume is 0.2233, so the estimates in table 3 suggest that a one-standard-deviation increase in trading volume predicts a 0.064-standard-deviation increase in Google searches, and a 0.068 standard deviation increase in Wikipedia Page views. These effects are significantly different from zero at the 10 percent level (p = 0.068 for Google, p = 0.055 for Wikipedia). We find that searches on date t are not significantly related to trading volume on date t+1 or beyond. All of the association of market movement

[20] Note that these transactions include many where the buyer and/or seller is not subject to capital gains taxes, may not be a human trader capable of information search, or where the asset does not have an accrued gain that will be subject to taxation. These possibilities do not threaten our identification so long as the percentage of transactions that is not subject to capital gains taxes does not change substantially from day to day.

with information search occurs on the same date as the market movement, or one day before. As a result, we are not able to determine from these data whether individuals seek information on capital gains taxes primarily before or after they make a decision regarding the sale of a capital asset. If they seek information before making a decision, they do so less than a day in advance, as far as these data are able to tell us.

We are sensitive to the possibility that much of the variation in these results may be driven by behavior during the 2008 stock market crash, a period of extremely high stock price volatility and trading volume. Columns (4) and (5) of Table 3 provide the estimates of the same regression specification, but limiting the inclusion of market variables to the period from September 2008 to February 2009. Columns (6) and (7) provide the estimates of the same regression specification, instead excluding this extraordinary period. For both Google and Wikipedia data, the estimated effect during the extraordinary period is much larger, although imprecisely estimated. However, the estimated effects we have described survive when the extraordinary period is excluded from the sample.

8.2 Do taxpayers search for capital gains tax information before they trade?

In this section we pursue an alternative indicator of the taxpayer demand for capital gains tax information that is itself based on observed search volume. In particular, we use Google Trends to obtain a measure of the volume of searches for phrases related to personal investment advice such as "stock advice", "should buy stock," "should sell stock," and "investment advice." In so doing, we hope to learn more about the timing of the relationship between search for tax-related information and decisions about whether to buy, sell, or hold assets. Note that, although Google searches for investment advice can lead one to websites that purport to provide such advice, searches for "capital gains tax" generally do not.

Table 4 shows the results of including the investment advice measure as an explanatory variable in a regression on searches for capital gains tax information, both in addition to stock trading volume and as a replacement for it.[21] Several results of interest emerge.

First, consider the results for Google searches. When investment advice search volume is included as an alternative to stock trading volume, the same positive contemporaneous association appears, and the effect of a one standard deviation change on capital gains tax related searches is 40 percent lower. In a "horse race" when both stock trading volume and investment-advice search volume are included as explanatory variables, unsurprisingly the statistical significance of the former declines because of the high correlation between the two variables. In the horse race, investment advice search volume wins, retaining its significance while stock trading volume losses its own significance. Strikingly, the one-day lag *and lead* values of stock investment advice search volume are significantly positively associated with capital gains tax search volume. This estimated association is consistent with a story that taxpayers first recognize the need for information regarding buying or selling stock and, in the process, learn that relevant to this decision are the tax consequences.[22]

8.3 Google Correlate

While the above mentioned stock-related terms that we selected (e.g., "stock advice") were correlated with searches for capital gains taxes in predictable ways, Google, through its application Google Correlate, also has the ability to provide a list of search terms most highly correlated over time with any given user provided search term. Consistent with our other evidence that taxpayers simultaneously search for information about capital gains taxes and investing, among the top 20 search terms related to "capital gains tax" are "stock purchase,"

[21] Appendix Figure A.2 graphically depicts the lead-lag structure estimated when stock advice search volume is included in the absence of stock market trading volume.

[22] The key advantage of the stock investment advice search volume variable is that it likely captures the extent of taxpayer demand for information for which capital gains tax knowledge is crucial. One potential disadvantage is that in regression analysis one may pick up any shocks that affect all Google Searches. To that point note in Table 5 that the same pattern of results also applies when the volume of Wikipedia searches is the dependent variable, with one exception: the lead relationship, although positive, does not reach statistical significance.

"investing," "td waterhouse," "fidelity mutual funds," and "mutual funds."[23]

8.4 Capital Losses

Because the capital gains tax rules related to the sales of assets with capital losses are especially important for many tax minimization strategies, such as loss harvesting, we also construct two measures of information search related specifically to capital losses. The first is weekly Google searches for the phrase "capital loss," and the second is monthly searches related to capital losses using the search functionality on the IRS website, IRS.gov. We obtain the latter measure directly from the IRS.[24] These time series are plotted in Figure 5.

As with searches for capital gains, much of the variation is seasonal: people tend to search for information on capital losses during tax season. There is also typically an increase in searches in December of each year, when some taxpayers "harvest" capital losses to reduce their tax liability. Of particular interest is the fact that searches for information on capital losses increased dramatically during October 2008, and surged even further in December of 2008. When the crash began in October, investors began to research the tax implications of the unrealized or realized losses they had sustained, perhaps evaluating the merits of pulling their wealth out of (or investing in) the declining stock market. Furthermore, for an investor who had lost money in the crash, harvesting capital losses before the end of 2008 could reduce the taxpayer's tax liability substantially.[25] The heightened importance of loss harvesting from the stock market crash apparently caused even more information search at year-end.

For both Google and IRS searches, we can also see that searches for information on capital losses *during tax season* are higher after the 2008 crash than before, for

---

[23] http://www.google.com/trends/correlate/search?e=capital+gains+tax&t=weekly&filter=capital+gains+tax&p=us#default,30 displays the full results. We thank Hal Varian for suggesting the use of Google Correlate.

[24] The IRS provided to us, at a monthly level, the most frequent search terms, and the number of searches for all terms with the word "capital", and "loss", in the search.

[25] This would naturally require that the investor had capital gains to offset with these losses.

three years after the crash. For several years after the crash investors realized capital losses with greater frequency than before the crash. At the same time, information search for tax compliance purposes also increased after 2008. This constitutes additional evidence of spillovers between macroeconomic changes and information search for the purpose of tax compliance, discussed further in Section 8.4.

8.5 Delayed effects

Not all information searches will happen immediately before or following some event. In particular, behavior and market movement during the year can be associated with information search at year end, as individuals decide whether to realize capital gains or losses prior to a time notch. To examine this possibility, we include an interaction of a year-end dummy—equal to one in the last five days of the year and zero otherwise—with aggregated yearly trading volume (in Table 5, *End of Year X Log Vol. This Yr.*) and the buy-and-hold market return for the year that is ending (*End of Year X Mkt. Ret. This Yr.*). We obtain the buy-and-hold market return using the CRSP daily return (variable VWRETD), cumulated over the entire year. The yearly buy-and-hold return is a proxy for the overall market performance that year, although it does not measure exactly accrued net capital gains from the sale of stock.

Individuals may also increase search activity during tax season of the subsequent year, as they research the compliance consequences of their previous actions. We explore this possibility by including an interaction of a tax season dummy—equal to 1 between January 30 and April 15—with aggregated yearly trading volume (*Tax Season x Log Vol. Prvs. Yr.*) and stock market returns from the previous year (*Tax Season x Mkt. Ret. Prvs. Yr.*). This allows us to examine if high trading volume and/or returns in a year are associated with greater information search for compliance purposes in the following year. The estimation controls for weekly and yearly seasonality, and the impact of events (as in Table 3). We also include the yearly aggregates with which we interact end-of-year and tax-season dummies.

The results of this estimation are provided in Table 5. High trading volume in year *y* is associated with elevated information search through Google and Wikipedia at the end of year *y*, and with elevated information search through Google, Wikipedia, and the IRS telephone hotline during tax season of year *y*+1. High stock market returns in year *y* are associated with elevated information search through Google and Wikipedia at the end of year *y*, but not with information search during filing season. This pattern is consistent with market returns affecting the gain to making tax-efficient decisions at year end—because timing matters more for realizing large capital gains than small ones—while high trading volume affects both the gain to making tax efficient decisions at year end *and* the number of taxpayers who will need to understand capital gains to complete their tax return come tax season.

9 Lessons from Aggregate Data on IRS Website and Call Line Usage

Figure 6 plots time series data on the use of the IRS website and call-line usage that are not specific to capital gains taxes. In order for their effects to be visible in broader measures of information search, events must significantly change the importance of understanding taxes at a given time, for a large number of taxpayers. Unsurprisingly, the IRS.gov domain and call line experience elevated traffic during tax season. Note that the amount of traffic experienced during tax season increases over time for the web page, probably due to increased use of the Internet by taxpayers. The call line, in contrast, decreased in usage from 1999 to 2008, and then experienced a resurgence. Together these results suggest that the Internet has not completely crowded out the use of the call line.

Most interestingly, we observe an abnormal surge in visits to IRS web pages and calls line traffic in May of 2008. Search volume usually drops sharply after the mid-April tax deadline, but in 2008 it remained high throughout the month of May. Among the top five searches on the IRS web page during this month were "stimulus," "rebate," and "stimulus check," and many of the top pages viewed also dealt with the stimulus rebates. The tax rebates enacted by the Economic Stimulus Act of 2008 led millions of taxpayers to visit the IRS web page to investigate how the federal stimulus program affected them. From more detailed data on the use of the call line, we know that the IRS received over 50 million phone calls on stimulus checks, corresponding to at least 11 million unique taxpayers. This corroborates the evidence provided by Sahm, Shapiro, and Slemrod (2012) that these rebate checks were a salient form of economic stimulus with a relatively high (compared to a reduction in employer withholding rates) marginal propensity to consume. Using the intuition of rational attention, it makes more sense for individuals to seek information in May about their stimulus check if they intend to spend it compared to if they intend to save it.

10 Conclusions

It is well-established that in general taxpayers know little about the US income tax, and have systematic misperceptions. Given that acquiring information is costly, it may be optimal for individuals to learn only *if* the expected return is high enough and only *when* the information is most useful, known as rational attention. Because people are learning—and forgetting—things all the time, the process of net information acquisition is critical to a dynamic understanding of tax salience. Using newly available IRS administrative data and publicly available information on Google and Wikipedia searches, this paper establishes that people seek information about the US income tax in systematic ways that are consistent with the idea of rational attention. When policies change or seem likely to change, when filing deadlines or time notches loom, people turn to online resources like Google, Wikipedia and the IRS website, as well as traditional information resources like the IRS telephone hotline, to learn how the tax code affects them. In addition, people search for tax-related information when newsworthy events make taxes more salient, and in so doing they may incidentally obtain information relevant to their own decisions. When policy or news events generate exogenous shocks to the demand for information, the responsiveness of information search to the event occurs remarkably quickly: search behavior usually spikes on the same day as the event, and falls back to baseline within three or four days. This timing pattern suggests that Sims' (2003) mechanism for modeling rational attention more closely matches the data than the one proposed by Reis (2006).

We also present somewhat weaker evidence regarding to what extent acquiring information about taxes leads individuals to alter behavior. We show that individuals sought information about capital losses at year ends, especially after the stock market crash of 2008, since harvesting losses provides an opportunity to reduce one's tax liability. This finding is consistent with many investors not knowing the rules governing capital losses well enough to confidently apply them to make their behavior tax efficient in the wake of the crash. Information acquisition is a necessary component of the response to tax incentives, especially where more obscure details of the tax code are concerned.

Our attempt to learn about taxpayer information search is limited because we cannot account for all sources of information. For example, we cannot observe taxpayers obtaining information from paid professionals. About 60 percent of taxpayers do use a paid preparer when they file a tax return. Some of our measures may, however, capture financial advisors' information search. Further, we cannot rule out "learning by doing," whereby individuals learn about tax incentives while filling out a return and then improve their future decisions. Nor can we observe learning through social networks (as documented in Alstadsæter, Kopczuk, and Telle, 2010). As in all similar studies, we also cannot quantify the *amount* of learning that occurs via acquiring any one piece of information. Finally, taxpayers for whom capital gains taxes matter are generally wealthier than the typical taxpayer. These individuals may be more sophisticated in their use of information than the typical taxpayer, and many can afford to pay for high-quality information from financial advisors. Future work should examine patterns in information search about other policies for this reason.

References

Alstadsæter, A., W. Kopczuk, and K. Telle. (2010). Social Networks and Dynamics of Tax Avoidance: Evidence from a Well-defined Norwegian Tax Shelter. Working paper.

Blendon, R.J., S.R. Pelletier, M.D. Rosenbaum, and M. Brodie. (2003). "Tax Uncertainty." *Brookings Review* (Summer): 28-31.

Brown C.V. (1968). "Misconceptions about Income Tax and Incentives." *Scottish Journal of Political Economy* 16(2): 1-12.

Chang, E. C., and J.M. Pinegar. (1986). "Return Seasonality and Tax-loss Selling in the Market for Long-term Government and Corporate Bonds." *Journal of Financial Economics* 17(2): 391-415.

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

Connor, C.E., H.E. Egeth, and S. Yantis. (2004). "Visual Attention: Bottom-Up versus Top-Down." *Current Biology* 14: R850-R852.

Da, Z., J. Engelberg, and P. Gao. (2011). "In Search of Attention." *Journal of Finance* 66(5): 1461-1499.

de Bartoleme, C.A.M. (1995). "Which Tax Rate Do People Use: Average or Marginal?" *Journal of Public Economics* 56(1): 79-96.

Delli Carpini, M.X. and Keeter, S. (1996). *What Americans Know about Politics and Why It Matters*. New Haven, CT: Yale University Press.

Drake, M.S., D.T. Roulstone, and J.R. Thornock. (2012). "Investor Information Demand: Evidence from Google Searches around Earnings Announcements." *Journal of Accounting Research* 50(4): 1001-1040.

Fujii, E.T., and C.B. Hawley. (1988) "On the Accuracy of Tax Perceptions." *The Review of Economics and Statistics* 70(2): 344-347

Ivković, Z., J. Poterba, and S. Weisbenner (2005). "Tax-Motivated Trading by Individual Investors." *American Economic Review* 95(5): 1605-1630.

Liebman, J.B. and R.J. Zeckhauser. (2004). "Schmeduling." Unpublished.

Muth, J. F. (1961). "Rational Expectations and the Theory of Price Movements." *Econometrica* 29(3): 315-335.

Poterba, J.M. and S.J. Weisbenner. (2001). "Capital Gains Tax Rules, Tax Loss Trading, and Turn-of-the-Year Returns," *Journal of Finance* 56(1): 353-368.

Reis, R. (2006). "Inattentive Consumers." *Journal of Monetary Economics*, *53*(8), 1761-1800.

Sahm, C.R., M.D. Shapiro, and J.B. Slemrod. (2010). "Household Response to the

2008 Tax Rebate: Survey Evidence and Aggregate Implications." In *Tax Policy and the Economy, Volume 24*, edited by Jeffrey R. Brown, pp. 69-110. Chicago: Chicago University Press.

Sheffrin, S.M. (1994). "Perceptions of Fairness in the Crucible of Tax Policy." In *Tax Progressivity and Income Inequality*, edited by Joel Slemrod, pp. 309-334. Cambridge: Cambridge University Press.

Sims, C. A. (2003). "Implications of Rational Inattention." *Journal of Monetary Economics* 50(3): 665-690.

Slemrod, J. (1995). "The Compliance Cost of Income Taxation, and How to Reduce It," National Commission on Economic Growth and Tax Reform (Kemp Commission), Oral testimony given July 12, 1995.

Slemrod, J., Christian, C., London, R., and Parker, J. A. (1997). "April 15 Syndrome." *Economic Inquiry 35*(4): 695-709.

Slemrod, J. (2004). "New Estimates of the Compliance Cost of Income Taxation," Committee on Ways and Means, Subcommittee on Oversight, Hearing on Tax Simplification, invited written testimony given June 15, 2004.

Starks, L.T., L. Yong, and L. Zheng. (2006). "Tax-Loss Selling and the January Effect: Evidence from Municipal Bond Closed-End Funds." *Journal of Finance* 61(6): 3049-3067.

Theeuwes, J. (1991). "Exogenous and Endogenous Control of Attention: The Effect of Visual Onsets and Offsets." *Perception & Psychophysics* 49(1): 83-90.

TABLE 1: TESTABLE IMPLICATIONS OF THEORIES OF INFORMATION AND ATTENTION

Behavioral assumption	Should information search respond to… time notches?	macroeconomic change?	enactment of policy changes?	implementation of policy changes?	filing deadlines?	news events?
Full information	No	No	No	No	No	No
Rational expectations	No	No	Yes, instantly	No	No	No
Rational attention, Sims (2003)	Yes	Yes, instantly	Yes, instantly	Yes, instantly	Yes	No*
Rational attention, Reis (2006)	Yes	Yes, with delay**	Yes, with delay**	Yes, with delay**	Yes	No*
Exogenous Attention (Salience)	No	No	Yes, while in the news	Yes, while in the news	Yes	Yes
Summary of Our Findings	Yes: Dec 31 time notch, esp. capital losses (due to loss harvesting)	Yes, instantly: day-of stock market trading volume, response to stock market crash	Yes, instantly/while in the news: ATRA, presidential debates and elections	Yes: May 2008 stimulus	Yes: seasonality due to tax season	Yes, instantly: release of Mitt Romney's tax returns

Notes: *News events may inspire attention “rationally" due to a preference for being informed in public discussion of current events, but it will not inspire rational attention for the purpose of economic behavior, either capital gains realizations or tax compliance. ** By "with delay," we mean that the surge in information search as a result of the event should be spread out over a longer time period in the Reis (2006) model compared to the Sims (2003) model, as different taxpayers reach the date at which they (previously) decided to update at different times.

TABLE 2: ESTIMATING THE IMPACT OF EVENTS ON INFORMATION SEARCH FOR CAPITAL GAINS TAXES

	Google Searches	Wikipedia Page Views	Calls to IRS
Obama/McCain Debate	12.08	5.469	-0.222
	93.893	2662.510	-16.171
	(0.000)	(0.015)	(1.000)
Obama Elected	16.04	5.873	-0.132
	124.658	2914.756	-12.105
	(0.000)	(0.008)	(0.999)
Mitt Romney's 2010 tax return released	20.820	34.959	0.379
	161.000	17351.698	-559.084
	(0.000)	(0.000)	(0.000)
Mitt Romney's 2011 tax return released	20.720	12.647	-0.341
	55.689	6277.817	-32.471
	(0.004)	(0.000)	(0.999)
Obama/Romney Debate	9.39	9.880	0.251
	72.989	4904.106	-17.693
	(0.000)	(0.000)	(0.999)
Obama Re-Elected	9.14	14.563	-0.31
	71.023	7228.503	-14.267
	(0.000)	(0.000)	(1.000)
American Taxpayer Relief Act signed	15.310	9.623	-6.170
	118.978	4776.242	-188.522
	(0.000)	(0.000)	(0.000)
Sample Period	Jan 2, 2004— March 21, 2013	Dec 9, 2007— March 31, 2013	Jan 2, 2002— March 31, 2013
Number of Days	3364	1938	4100

Notes: The event dates listed as dependent variables represent, in order, Oct 15, 2008 (Presidential debate between Barack Obama and John McCain), November 4, 2008 (Barack Obama elected President), January 24, 2012 (Mitt Romney releases his 2010 tax return), September 21, 2012(Mitt Romney releases his 2011 tax return), Oct 16, 2012 (Presidential debate between Barack Obama and Mitt Romney), Nov 6, 2012 (Barack Obama re-Elected), Jan 2, 2013 (American Taxpayer Relief Act signed into law, includes an increase in top capital gains tax rate).For each event, the top number reports the cumulative information search attributed to this event, in daily standard deviation units, added over the five-day event window. The second number reports the same estimate in the original units of the search volume measure, i.e. the Google Trends index or, Wikipedia page views, or number of calls to the IRS. The bottom number, in parenthesis, reports the p-value from an F-test in which the null hypothesis is that the event had no impact on information search, i.e. that the variation in information searches over the event window is purely random. The estimation controls for variation due to yearly and weekly seasonality.

TABLE 3. INFORMATION SEARCH AND STOCK MARKET ACTIVITY

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample Period	Jan 2, 2004— Dec 29, 2012	Dec 9, 2007— Dec 29, 2012	Jan 7, 2002— Dec 29, 2012	Sep 1, 2008— Feb 28, 2009	Sep 1, 2008— Feb 28, 2009	Jan 2, 2004— Dec 29, 2012, except Sep 1, 2008— Feb 28, 2009	Dec 9, 2007— Dec 29, 2012, except Sep 1, 2008— Feb 28, 2009
Dependent Variable:	Google Searches (Standardized)	Wikipedia Page Views (Standardized)	Calls to IRS (Standardized)	Google Searches (Standardized)	Wikipedia Page Views (Standardized)	Google Searches (Standardized)	Wikipedia Page Views (Standardized)
Log Trading Volume	0.286	0.306	0.043	1.071	0.926	0.215	0.386
	(0.156)*	(0.159)*	(0.075)	(0.780)	(0.470)*	(0.159)	(0.171)**
Log Trading Volume Lag 1	-0.051	-0.072	0.099	-0.141	0.263	-0.11	-0.093
	(0.146)	(0.151)	(0.070)	(0.708)	(0.415)	(0.149)	(0.162)
Log Trading Volume Lead 1	0.062	-0.023	-0.037	-0.125	-0.123	0.11	-0.05
	(0.177)	(0.181)	(0.085)	(0.887)	(0.553)	(0.179)	(0.193)
Number of Days	2252	1266	2753	124	124	2128	1142

Notes: The dependent variable and log trading volume are de-trended prior to estimation, and the dependent variable is standardized by dividing by the standard deviation of the detrended data. We control for weekly and yearly seasonality and the events in Table 2 (except those events that occur outside the sample period, since market data are not available past December 29, 2012). Weekends and holidays are omitted, since stock market data are not generated on weekends. The regression included five market-dated leads and lags of each measure. We only report one lead and one lag for brevity and clarity (See Appendix Figure A.1). The coefficients are similar if we use 14 leads and lags instead of 5, and none of the results change substantially with the inclusion of lags of the dependent variable. Standard errors are provided in parentheses below point estimates. * indicates $p<0.10$, ** indicates $p<0.05$, and *** indicates $p<0.01$. The standard deviation of log trading volume is 0.223.

TABLE 4: SEARCHES FOR STOCK ADVICE AND SEARCHES FOR CAPITAL GAINS TAX

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample Period	Jan 2, 2004— Dec 29, 2012	Jan 2, 2004— Dec 29, 2012	Jan 7, 2004— Dec 29, 2012	Jan 7, 2004— Mar 21, 2013	Dec 9, 2007— Dec 29, 2012	Dec 9, 2007— Dec 29, 2012	Dec 9, 2007— Dec 29, 2012	Dec 9, 2007— Mar 31, 2013
Dependent Variable:	Google Searches (Standardized)				Wikipedia Page Views (Standardized)			
Log Trading Volume	0.286	0.174	0.130		0.306	0.162	0.178	
	(0.156)*	(0.155)	(0.155)		(0.159)*	(0.157)	(0.157)	
Log Trading Volume Lag 1	-0.051	-0.100	-0.06		-0.072	-0.141	-0.137	
	(0.146)	(0.145)	(0.146)		(0.151)	(0.147)	(0.148)	
Log Trading Volume Lead 1	0.062	0.023	0.002		-0.023	-0.056	-0.098	
	(0.177)	(0.175)	(0.175)		(0.181)	(0.177)	(0.178)	
Stock Advice Search		0.031	0.008	0.009		0.030	0.016	0.017
		(0.004)***	(0.005)	(0.005)*		(0.004)***	(0.006)***	(0.005)***
Stock Advice Search Lag 1			0.017	0.017			0.017	0.015
			(0.006)***	(0.006)***			(0.008)**	(0.008)*
Stock Advice Search Lead			0.009	0.008			0.005	0.003
			(0.005)**	(0.005)*			(0.006)	(0.006)
Number of Days	2252	2239	2200	2211	1266	1266	1266	1272

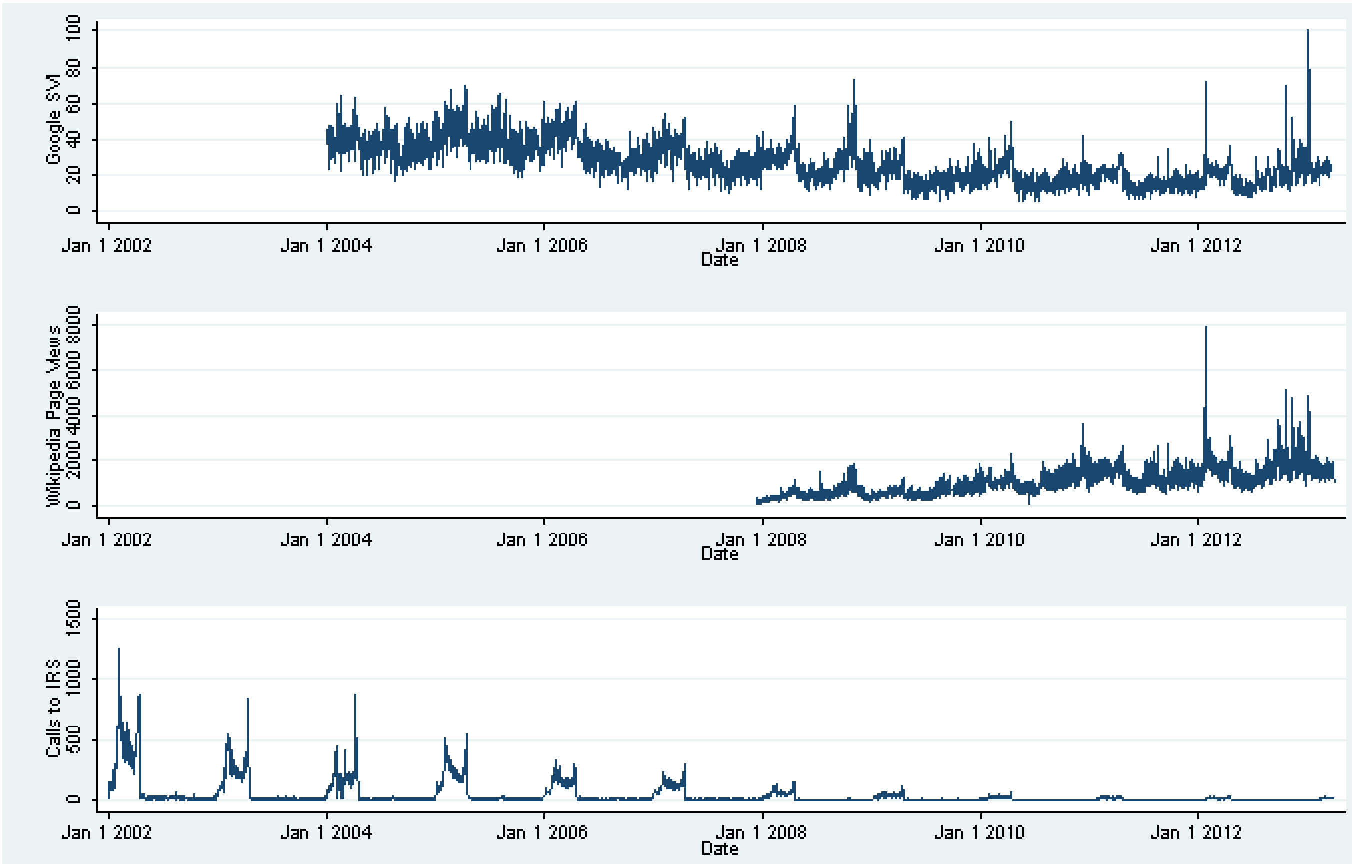
Notes: The dependent variable, log trading volume, and stock advice search volume are de-trended prior to estimation, and the dependent variable is standardized by dividing by the standard deviation of the detrended data. We control for weekly and yearly seasonality and the events in Table 1 (except those events that occur outside the sample period, since market data are not available past December 29, 2012). Weekends and holidays are omitted, since stock market data are not generated on weekends. The regression included five market-dated leads and lags of trading volume and stock advice search volume. We only report one lead and one lag for brevity and clarity (See Appendix Figure A.2). Standard errors are provided in parentheses below point estimates. * indicates p<0.10, ** indicates p<0.05, and *** indicates p<0.01. The standard deviation of log trading volume is 0.223. The standard deviation of Stock advice search volume is 4.458.

TABLE 5: DELAYED EFFECTS OF SEARCH ON MARKET OUTCOMES

	Google Searches (Standardized)	Wikipedia Page Views (Standardized)	Calls to IRS (Standardized)
Log Trading Volume, This Year	0.083	-0.169	-0.005
	(0.136)	(0.150)	(0.050)
Log Trading Volume, Previous Year	-0.039	0.263	-0.011
	(0.107)	(0.268)	(0.046)
End of Year *x* Log Volume This Year	0.029	0.036	-0.004
	(0.006)***	(0.006)***	(0.003)
Tax Season *x* Log Volume Previous Year	0.028	0.001	0.067
	(0.007)***	(0.007)	(0.003)***
Market Return, This Year	-0.161	-0.227	-0.015
	(0.111)	(0.206)	(0.037)
Market Return, Previous Year	0.094	0.01	0.024
	(0.100)	(0.121)	(0.248)
End of Year *x* Market Return This Year	1.502	1.975	0.238
	(0.606)**	(0.514)***	(0.248)
Tax Season *x* Market Return Previous Year	0.149	0.467	-0.142
	(0.165)	(0.156)***	(0.072)
Sample Period	Jan 2, 2004— Dec 31, 2012	Dec 9, 2007— Dec 31, 2012	Jan 7, 2002— Dec 31, 2012
Number of Days	3025	1695	3684

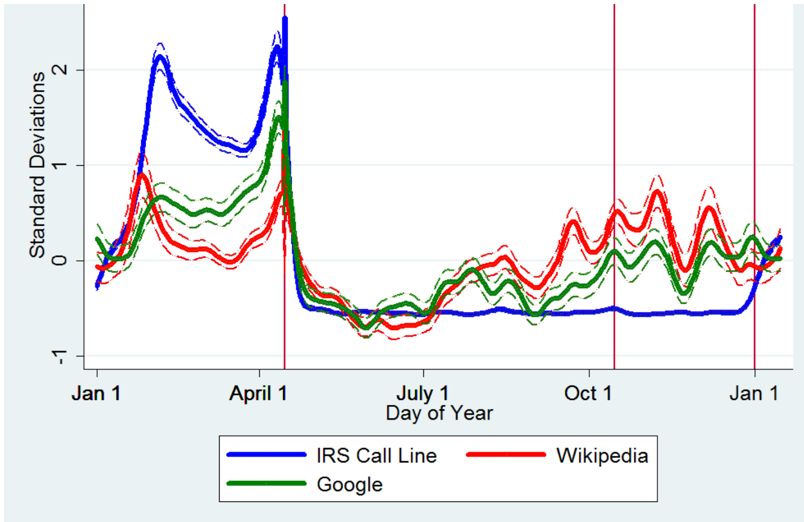
Notes: Data are de-trended prior to estimation, and the dependent variable is standardized by dividing by the standard deviation. We control for weekly and yearly seasonality and the events in Table 2 (except those events that occur outside the sample period, since market data are not available past Apr 8, 2012). Standard errors are provided in parentheses below point estimates. * indicates p<0.10, ** indicates p<0.05, and *** indicates p<0.01.

FIGURE 1: PLOTS OF THE RAW DATA



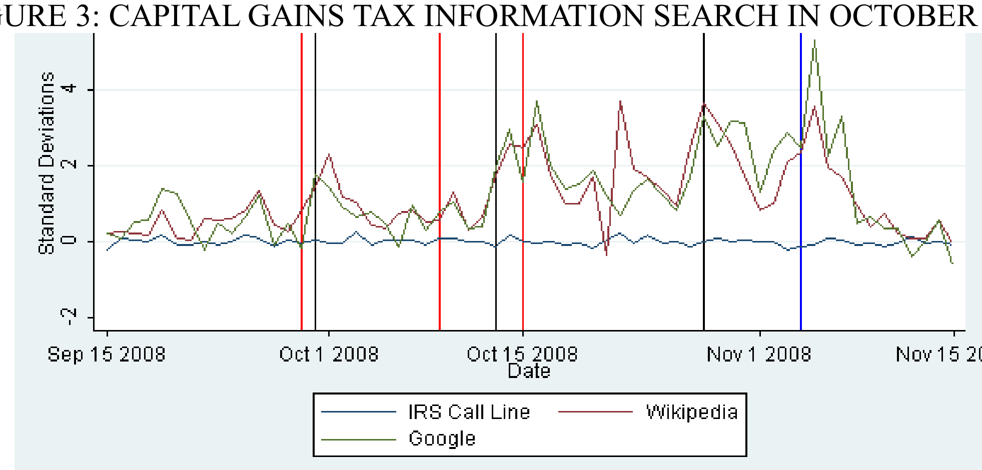
Notes: These graphs show the raw data for (in order) Google search volume, Wikipedia page views, and calls made to the IRS.

FIGURE 2: YEARLY SEASONALITY



Notes: Yearly seasonality is estimated using the smooth-fixed-effects method described in Section 5, with controls for the events in Table 2 and the market movements in Table 4. The dashed lines are the bounds of 95 percent confidence intervals. The three vertical lines correspond to April 15, October 15, and January 1, respectively.

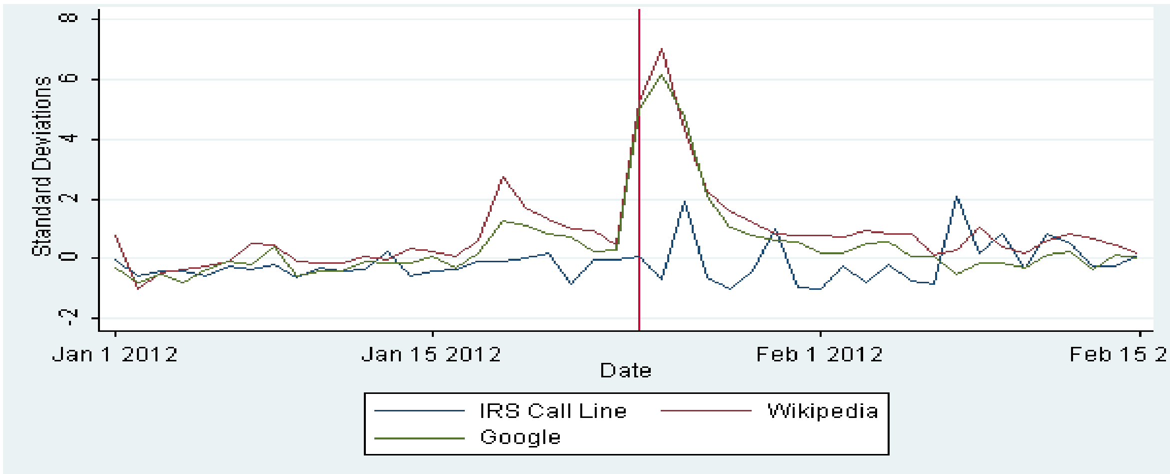
FIGURE 3: CAPITAL GAINS TAX INFORMATION SEARCH IN OCTOBER 2008



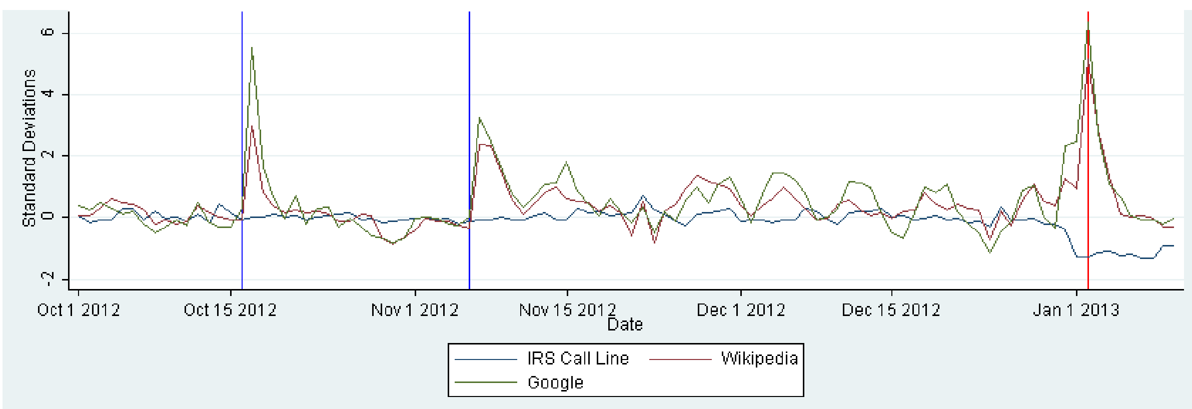
Notes: We plot the detrended, standardized, and seasonally adjusted data over a narrower time period to examine abnormal behavior in late 2008. The vertical red lines mark the worst three days of the stock market crash of 2008 according to the drop in the S&P 500 index: October 15 (-9.03%) , September 29 (-8.81%), and October 9 (-7.62%). The black lines mark the three largest gains in the S&P 500 during this volatile period: October 13 (11.58%), October 28 (10.79%) and September 30 (5.42%). October 15, 2008 was also the date of a Presidential Debate between Barack Obama and John McCain, in the course of which both candidates made proposals for changing capital gains tax rates The other debates occurred on September 26 and October 7, and did not discuss capital gains taxes. The vertical blue line marks the 2008 Presidential election, November 4, 2008

FIGURE 4: NARROW TIME WINDOWS OF SEARCH FOR TAX INFORMATION

Panel A. Capital Gains Tax Information Search in January 2012

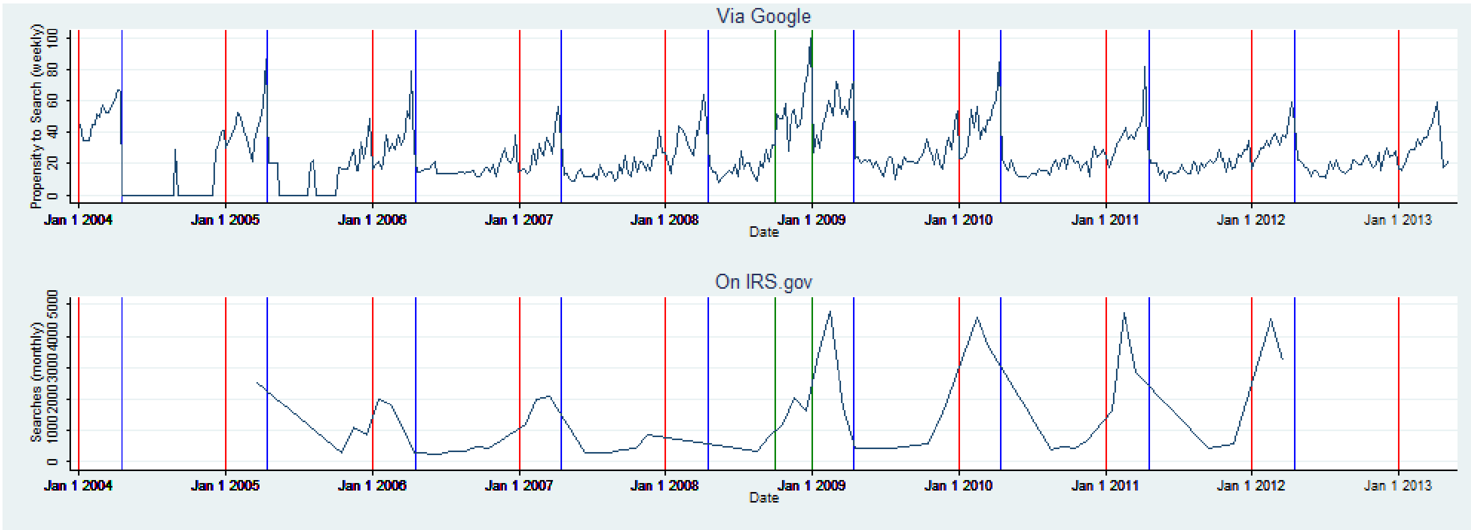


Panel B. Capital Gains Tax Information Search in November 2012 – January 2013



Notes: In Panel A, we plot the detrended, standardized, and seasonally adjusted data (from Figure 4) over a narrower time period to examine the spike in January 2012. The vertical red line marks January 24, 2012, the date that Mitt Romney released his 2010 tax return. The spike a few days prior is coincident with a related news story wherein Mr. Romney told the press about his effective tax rate. In Panel B, the first blue line marks the date of the Presidential debate on October 16, 2012, during which Barack Obama and Mitt Romney debated the merits of Romney's proposals for lowering capital gains tax rates. The second blue line marks the date of the Presidential election, November 7, 2013. The red vertical line marks the passage of the American Taxpayer Relief Act on January 2, 2013.

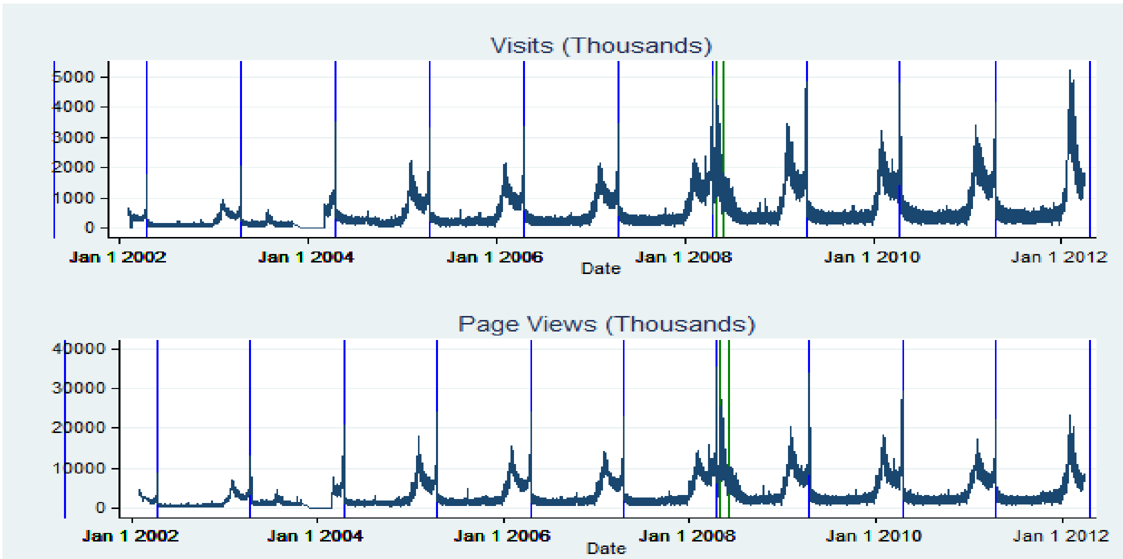
FIGURE 5: INFORMATION SEARCH ON CAPITAL LOSSES



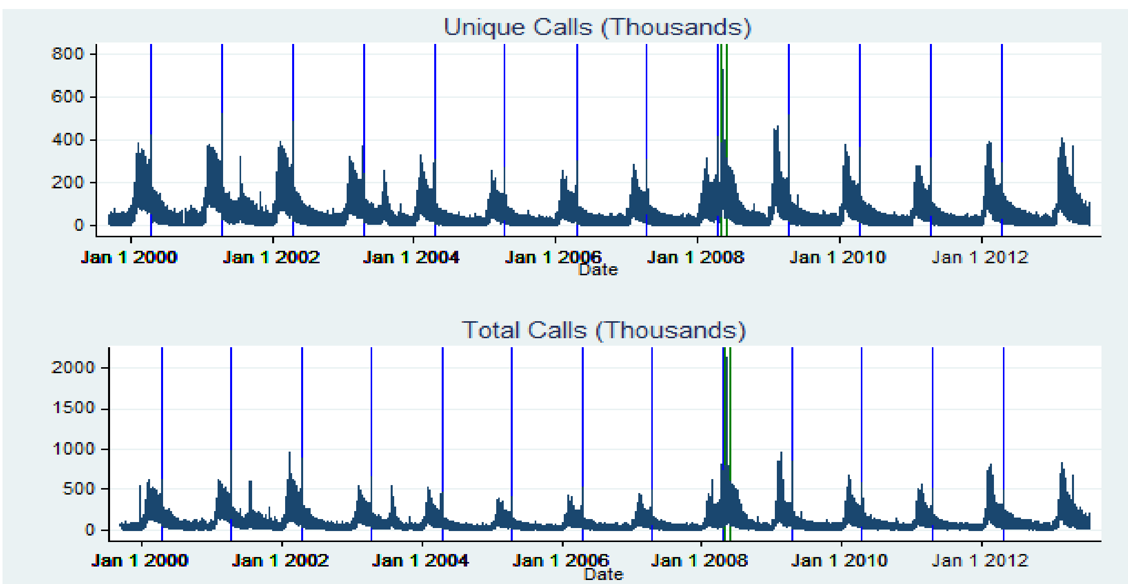
Notes: The first panel plots weekly Google Trends data on searches for just "capital loss." Google Trends data are missing for some dates in 2004 and 2005, when search volume was too low and Google does not provide that data. The second graph plots monthly searches in the search bar on the IRS home page for terms related to capital losses. The green lines correspond to October 1 and December 2008, to delineate the period in which there was increased information search on capital losses due to the 2008 stock market crash. Red lines denote January 1st of each year, to highlight that a spike in search volume typically occurs at the very end of each year. Blue lines denote mid-April filing deadlines, to highlight the spike occurring in the run-up to the filing deadline.

FIGURE 6: OVERALL USE OF IRS INFORMATIONAL RESOURCES

Panel A: Visits and Page Views of all URLs on the IRS.gov Domain.



Panel B: Calls to IRS Line for Individuals



Notes: The green vertical lines delineate the month of May, 2008. Blue vertical lines correspond to the mid-April filing deadline for each year. Panel A: Data are unavailable from November 1, 2003 to February 29, 2004. Panel B: The phone number for this line is 1-800-829-1040, a number made available numerous places on the IRS website and on the instructions for individual tax returns (form 1040). Unique callers are identified by the telephone number from which the call originates.

For Online Publication

Obtaining and Analyzing Google Trends Data

When querying Google Trends, the user provides 1) search terms, 2) a geographical window, and 3) a time range. As we are studying a tax issue within the United States, for all data used in this paper, the geographical range is specified to be the United States.

When the number of overall searches for a given term is too low, Google Trends will report an SVI of zero, or will report daily at a weekly or monthly level (as opposed to daily SVI data). We encounter this issue if we query searches for "capital gains tax" alone. As such, we use a set of search terms to maximize our sample period for which we are able to obtain daily data. The search terms we include are, according to Google Trends itself, highly correlated with searches for "capital gains tax." The set of search terms is the following:

- Capital gains tax
- Capital gains tax rate
- Capital gains taxes
- Capital gains tax rate
- Capital gains calculator
- Capital gains
- Capital gains rate

We have verified that 1) the daily time series of SVI for simply "capital gains tax" is virtually identical to the one from the broader set of search terms, but with fewer missing data, and 2) we obtain nearly identical results for event studies and market movement effects if we use simply searches for "capital gains tax" as our left-hand-side variable instead of the broader set of search terms, but with slightly larger standard errors (reflecting the decreased number of observations).

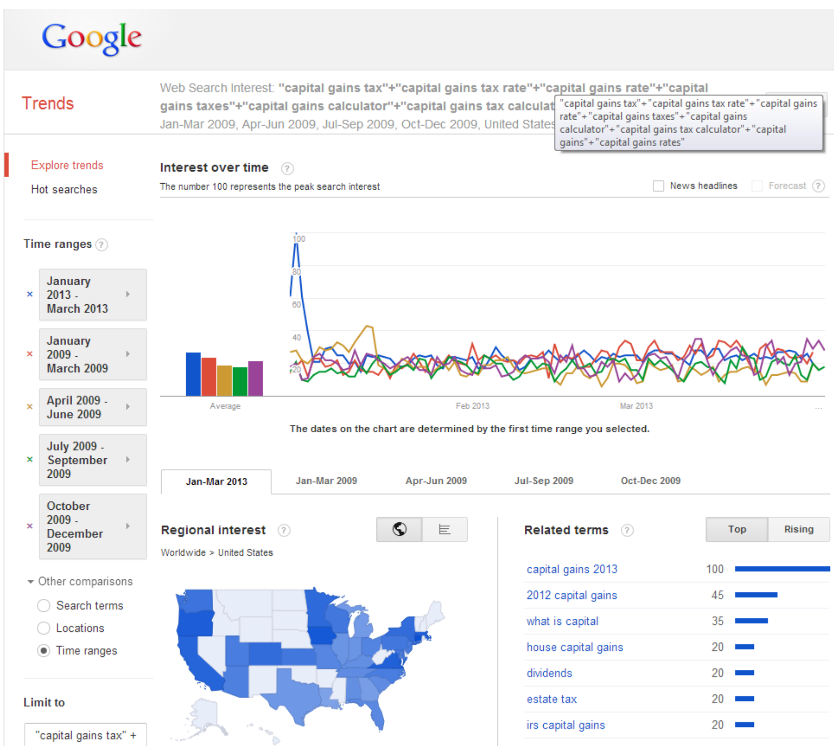
For the same reasons as above, we use multiple search terms related to stock advice. These are:

- Stock advice
- Stock market advice
- Stock tips
- Stocks to buy
- Stock to buy
- Stocks to sell
- Should buy stock
- Should sell stock
- Investment advice
- Investment tips

By default Google Trends provides weekly data when a user downloads a time series longer than three months. In order to access daily search volume data, one must query Google Trends in three month intervals. Fortunately, one can query several (up to five) three-month periods at once.

In order to obtain daily data while maintaining proper scaling of the variable across the entire time series, however, one must query Google Trends very carefully. The data are scaled so that SVI takes the value of 100 on the day with the highest propensity to search out of any date range and search terms. For example, if January 2, 2013 were the day with the highest propensity to search for capital gains tax terms (as it is), SVI would equal 100 on that day if January 2, 2013 were in the period provided by the user. In order to obtain daily data that is properly scaled for the full sample period, one must first find the single day with the highest search volume, and then include a time period containing that day along with sets of other three-month periods.

To get a properly scaled daily time series, we therefore include the time period January 1, 2013-March 31, 2013 in every single one of our queries for capital gains tax SVI, along with other three-month periods, until we obtain data for our entire sample period. The figure below shows an example of what such a query would look like to pull daily data for the year 2009. One can then download the data directly from this web page by clicking on the cog icon.



Estimation Procedure

Recall that the equation to be estimated is

$$I_{it} = \sum_k \left[\beta_{k0}F_{kt} + \beta_{k1}F_{k(t-1)} + \cdots + \beta_{k4}F_{k(t-4)}\right] + x_t'\gamma + f_i(DoY_t) + \delta_{DoW,i} + u_{it} \quad (1)$$

In order to consistently estimate equation (1), which contains a non-linear function and a set of linear covariates, we use the double residual regression method suggested by Robinson (1988), also discussed in Hardle and Linton (1994). This first step of this three-step estimator consists of several non-parametric regressions of the following form:

$$E[X_t] = f_i(DoY_t),$$

where X_t is the dependent variable or one of the linear covariates, i.e. I_{it}, F_{kt}, x_t or $\delta_{DoW,t}$. We then obtain the residuals from this regression, which we denote with star superscripts. The residuals represent the component of I_{it}, F_{kt}, x_t or $\delta_{DoW,t}$ not correlated with the general within-year pattern represented by $f_i(DoY_t)$. The second step estimates the linear components of the model consistently using ordinary-least-squares regression on these residuals:

$$E[I^*{}_{it}] = \sum_k \left[\beta_{k0}F_{kt}^* + \beta_{k1}F_{k(t-1)}^* + \cdots + \beta_{k4}F_{k(t-4)}^*\right] + x_t'^*\gamma + \delta_{DoW,i}^* \quad (2)$$

The third step regresses the residuals from the estimation of equation 2—denoted by $I^{**}{}_{it}$ — non-parametrically on DoY_t to obtain a consistent estimate of the function $f_i(DoY_t)$.

$$E[I^{**}{}_{it}] = f_i(DoY_t) \quad (3)$$

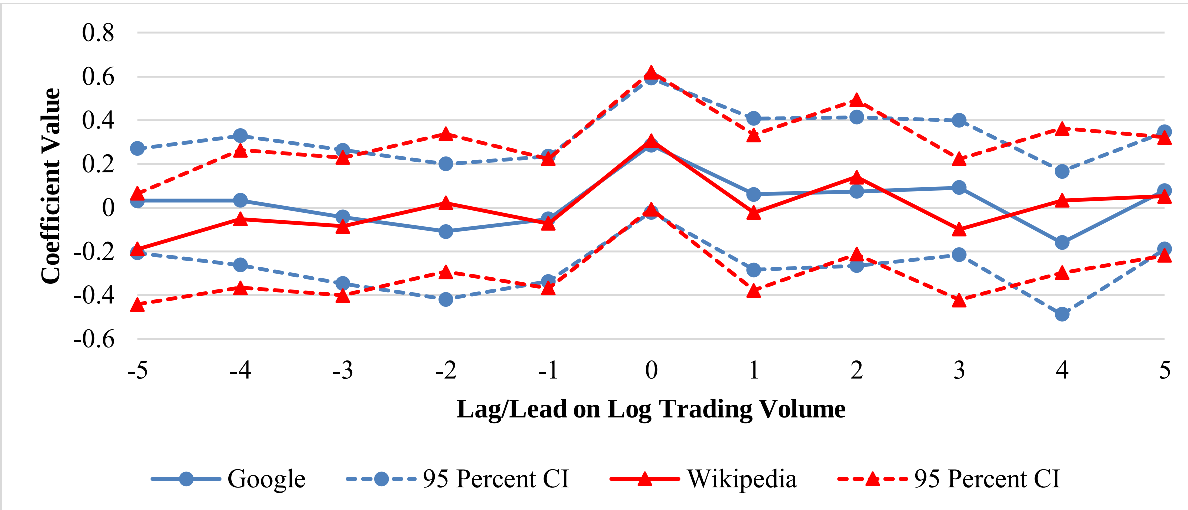
The first and third steps involve non-parametric regression, which Robinson (1988) suggests implementing via kernel-weighted local polynomial regression. In this paper, we use local linear regressions with a Gaussian kernel density function and a bandwidth of four days.[26] This procedure estimates a weighted OLS regression of the dependent variable on *DoY* at each value of *DoY* in the data, where the weights are determined by the kernel density function and the bandwidth. The bandwidth was selected to visually match the fixed-effects estimator of the function $f_i(DoY_t)$, but a data-driven choice of bandwidth—specifically selecting the bandwidth that minimizes the conditional weighted mean integrated squared error—yields nearly identical results. The estimates we present are also virtually unchanged by varying the

[26] This bandwidth and kernel density function applies to non-parametric estimations from the third stage of the procedure, and nonparametric estimations involving continuous variables in the first stage. For discrete variables in the first stage, such as event dummy variables, we use a bandwidth of zero, to reflect that there should be no smoothing at this stage. Another complication is that traditional statistical packages and programs will implement the smoothing on a linear variable, while day-of-the-year is a cyclical variable. That is, January 1 and December 31 should be adjacent for smoothing purposes. Ignoring this problem results in a discontinuity in the seasonal pattern between these two days. We eliminate the discontinuity by estimating the seasonal pattern twice: once where the discontinuity is imposed at January 1, and a second time where the discontinuity is imposed at the 200th day of the year (July 19th). Then, we replace the 10 days around January 1 from the first estimation with these days from the second estimation.

degree of the local polynomial, the bandwidth, and/or our choice of kernel function, with the exception that a wider bandwidth results in a smoother function that no longer resembles the fixed-effects estimates and a narrower bandwidth results in a more jagged function.

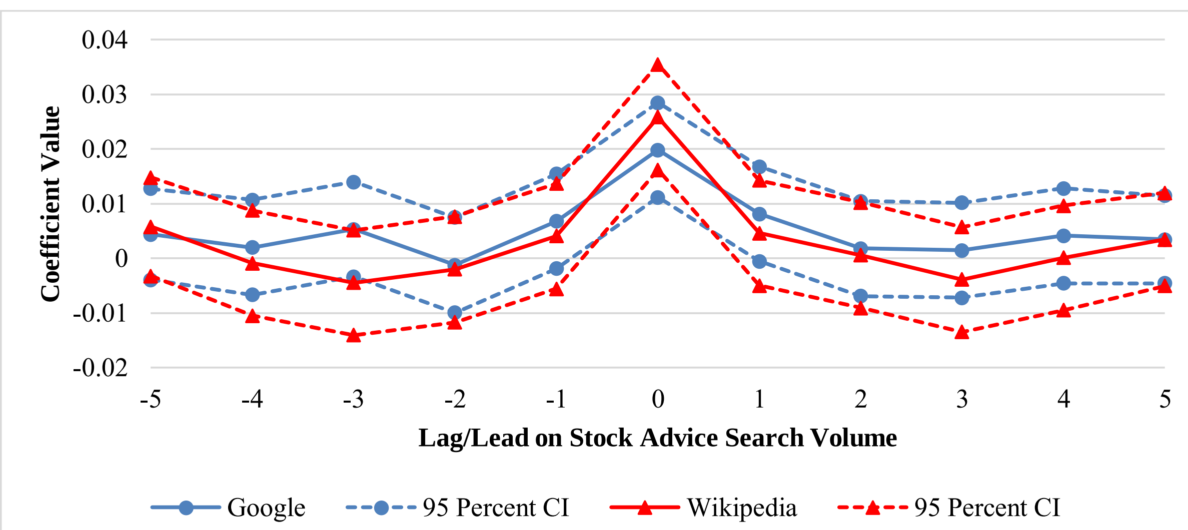
Supplementary Figures

FIGURE A.1 ESTIMATED COEFFICIENTS ON LOG TRADING VOLUME



Notes: Plotted are coefficients and confidence intervals from the regression of leads and lags of log trading volume on information search for capital gains tax. See also Table 3, Columns 1 and 2.

FIGURE A.2 ESTIMATED COEFICIENTS ON STOCK ADVICE SEARCH VOLUME



Notes: Plotted are coefficients and confidence intervals from the regression of leads and lags of stock advice Google search volume on information search for capital gains tax. See also Table 4, Columns 4 and 8.

Appendix References

Robinson, P.M. (1998). "Root-*N*-Consistent Semiparametric Regression." *Econometrica* 56(4): 931-954.

Hardle, W. and O. Linton. (1994). "Applied Nonparametric Methods." In *Handbook of Econometrics Volume IV,* edited by R.F. Engle and D.L. McFadden, pp. 2295-2239. Elsevier.