Investors' Information Choice

Astaiza-Gómez, José Gabriel

2021

Investors' Information Choice

Working Paper

José Gabriel Astaiza-Gómez[*]

October 4, 2021

Abstract: I estimate a demand model for online services of financial data, from a random parameters or mixed logit model, using a sample with searches at Bloomberg Terminals and at the EDGAR system. My preliminary results suggest that the substitution investors make of financial information providers, are affected by the subscription prices, investors' expectations on stock returns, and investors' income.

Key words: random parameters, open access services, subscription providers, market shares.

[*]I thank Karoll Gómez and Jorge Florez for their thorough readings and commentaries on my reports. All errors are mine.

1 Introduction

Investors need to get informed in order to take good investment decisions, and the idea that having many providers of information, improves the decision making process of investors, is intuitive. Nevertheless, as acquiring data is costly and attention is a scarce cognitive resource (Peng and Xiong, 2006), investors also need to take decisions on which source will provide them the needed information. The current massive use of the internet offers to investors a large number of sources which include no-fee open access providers such as Yahoo Finance and EDGAR, that co-exist with the services of subscription providers such as Bloomberg, Refinitiv Eikon, FactSet and Capital IQ. These sources are not perfect substitutes: although the sets of information they offer have common elements, some may include additional information or present it in a different manner. In this research, I empirically study the drivers of investors' demand for financial information services.

In previous studies, little attention has been devoted to analyzing investors' choices over information sources. While the study of product demand is extensive in the academic literature, understanding investors substitution patterns over information providers has its own issues. First, financial information about stock markets is useful for investors as long as they have the financial capability to buy financial assets, and therefore, even with open-access information providers, income is important to understand information acquisition activities. Second, financial information is not consumed for its own sake but it is acquired with the purpose of making a decision on another product: a financial asset. As the acquisition of financial information is expected to help in the decision making process of asset allocation, it requires the investor to process it, making it relevant to evaluate the potential effects of investor's personal characteristics, such as financial literacy and willingness to take risks, on changes in choices over information providers.

I estimate a demand model for financial information services, using data on investor demographics, stocks and search activities. In particular, I use survey results at the investor level from the FINRA Foundation, returns and fundamental variables of stocks belonging to the CRSP index, daily non-robot EDGAR downloads and the News Heat index of Bloomberg searches. I estimate the parameters of a random parameters or mixed logit model, controlling for unobserved service characteristics as well as for unobserved investor characteristics, as proposed in Nevo (2001). As the service price is endogenous, I use, as instruments, (lagged) Google's SVI of the expression "Bloomberg Terminal" as well as sales revenues of Microsoft, arguing that the popularity of Bloomberg Terminals is related to Bloomberg's capacity to charge higher prices, and based on the facts that the Terminals are not available for MAC operating systems and many functionalities involve an intensive use of Excel. In addition, as the demand for information may influence the expectations investors have on stock returns, I use as instruments a set of variables that I argue, are not affected by choices on information services, such as firm return on assets or earnings per share volatility, among others. I find that subscription prices as well as expected returns are important to explain financial information demand, and that income have an influence on investor's marginal valuation of expected returns when deciding over sources of information. This research is, to the best of my knowledge, the first attempt to understand the substitution patterns between sources of information.

While, as discussed in section 7, I can improve my estimates, my current results show that there are investor characteristics that induce investors to change their information providers, and that some investors perceive, to some extent, service informational advantages. This is important because new open-access sources of financial information may affect the viability for existent subscription providers, who in turn may react in order to stay in the market, and thus, investors' decisions on services affect the overall informational environment of the stock market and investors well being. For instance, a subscription provider may reduce its effort to deliver that information also available on open access providers or may increase its effort to present this same information in a more friendly manner or in higher detail.

There are eight sections in this document including the introduction. In section two I present the literature related to financial information demand. In section three I show my sample, describing the data related to information acquisition activities, as well as that related to investor demographics. In sections four and five I explain my model of investor's choices over sources of financial information, the definition of the outside option and identification issues. In section six I show my results, and in sections seven and eight, I discuss my results and conclude.

2 Literature Review

The relevance of financial information demand is well known in the theoretical literature. For instance, in a model where acquiring information is costly, Grossman and Stiglitz (1980) show that the degree in which stock prices convey information from the informed to the uninformed investors, depends on the number of individuals who get informed. At the same time, the availability of many information sources require investors to allocate their attention, which is a limited resource. From the experimental and theoretical literature, it is clear that information search is a costly cognitive operation, for which individuals must also make attention allocation decisions (Gabaix, et al., 2006; Sadler, 2021). Recent empirical literature in finance (e.g. Hu, et al., 2021) recognize that investors have limited attention and expose that the improvement of the informational environment for investors, that comes from having a greater number of information channels is not straightforward, since having a great number of them creates a poverty of attention, and thus, there is a need to make an optimal allocation of attention, across the overabundant sources that might consume it (Da, Engelberg and Gao, 2011). The implications of attention allocation decisions for stock markets has been studied in the theoretical literature, as in Peng and Xiong (2006), who model a representative investor who must allocate his attention among types of information, namely, market information, sector information and firm information. In my research, I empirically study investors' choice of information providers.

The demand for information is likely related to expectations on stock returns. For instance, Willinger (1989) models an investor who solves a problem of asset allocation in which he must also decide to pay (or not) the cost of acquiring information. As the contents of the new information are uncertain ex-ante, information itself is risky. In Willinger's (1989) model, information increases the expected return of the investor's portfolio and thus is valuable. Moreover, Vlastakis and Markellos (2012) empirically show that Google's search index on company names, increases for periods of higher stock returns. Information demand also likely depends on some investor characteristics such as financial literacy and the willingness to take risks. Willinger (1989), finds that the expected value of information (first defined in LaValle, 1968) and risk aversion are related. Lipman (1991) posits that allocation decisions (here choosing information providers) vary across individuals, even if we know our preferences and our feasible sets, since we may not know, all the logical implications of a same fact. For example, it is very likely that one does not know many theorems of set theory, even when these are logically implied by the axioms one knows. A more interesting example in the context of this paper, can be found in the results of the FINRA Foundation's survey of 2015, in which around 16% of those who had investments in stocks, gave an answer different to "More than $102" to the following question: "Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow?."

Although, to the best of my knowledge, the empirical literature on investors' choices of financial information sources is scarce, there is a considerable amount of literature studying the characteristics of the users of a single open-access provider of information. Da, Engelberg and Gao (2011) use the Search Volume Index (SVI) of Google and find that increments in SVI, are related to trading of less sophisticated retail investors. Also, Behrendt, Peter and Zimmermann (2020) show that patterns of retail (collective) trading, are related to Wikipedia searches for firm information. Loughran and McDonald (2017), find that the most requested EDGAR filings of non-robot investors, tend to be about popular companies such as Facebook, and argue that retail investors would not be using robots to search information at the SEC.gov website. Chi and Shanthikumar (2018) find that retail trading (buying and selling) is related to searches for 10-K and 10-Q filings at EDGAR and Asthana, Balsam and Sankaraguruswamy (2004) show that when firms filed the form 10-K on EDGAR for the first time there was an increase in the volume of small trades whereas there is not an effect for large investors. At the same time, subscription providers are more likely to be used by institutional investors. For instance, Ben-Rephael, Da, and Israelsen (2017) propose a measure of institutional investor attention based upon searches and reading activity at Bloomberg terminals. The authors document that, as of August 26, 2016, around 80% of Blommberg Terminal users worked in financial industries including banking, asset management, and institutional financial services, with 32% of the job titles being portfolio managers or traders, 19% presidents or directors, and 17% being analysts, including buy-side and sell-side ones. This is in line with the research that, following the intuition, assume that Bloomberg is a source of information for more sophisticated or institutional investors, which include Da, Engelberg and Gao (2011) and Wang (2020).

3 Data and Preliminary Evidence

This paper uses four main data sets. First, it uses data on search activities at two different financial information services, namely, EDGAR and Bloomberg. Second, it uses data on investor characteristics, or demographics, from the Financial Industry Regulatory Authority (FINRA) Foundation, and in particular, it uses the results of the National Financial Capability Study of 2012 and 2015. Third, data on stocks belonging to the CRSP index (3,611) are used in this paper. I collect historical stock prices (from which I calculate stock returns), earnings per share, return on assets, market capitalization, number of analyst recommendations, target prices (from which I calculate bias volatility), the sectors to which they belong and whether the stock is in the list of section 13(f) securities of the SEC. Finally, this research uses historical data on Google's SVI for the expression "Bloomberg Terminal", as well as Microsoft's sales revenues. I provide more details in the next subsections.

3.1 Financial Information Services

For firms included in the Center for Research in Security Prices (CRSP) stock index from the second quarter of 2010 to the fourth quarter of 2015, I observe daily data on the News Heat - Daily Max Readership index of Bloomberg. The index is constructed by Bloomberg based upon the "number of times each article is read by its users, as well as the number of times users search for news for a specific stock" (Ben-Rephael, Da and Israelsen, 2017) and takes higher values for higher levels of readers activity going from 0 to 4. Additionally, I observe the daily non-robot searches of EDGAR filings through SEC.gov. The Division of Economic and Risk Analysis (DERA) constructed the EDGAR log file data set containing statistics on user access to the SEC.gov website and is "intended to provide insight into the usage of publicly accessible EDGAR company filings".

Among other variables, the EDGAR log file data, includes the IP addresses, dates, Central Index Key (CIK) codes and whether the user self-identified as a crawler. In the SEC website[1] there are 2,880 zip folders in the EDGAR log file data set for the period 2010 - 2017, one for each day, and each folder contains a "README" file documenting the variables, and a *csv* file with the statistics on SEC.gov website traffic, which can include more than ten million entries (e.g. the file of January first of 2015 contains 15'682,916 rows and 15 columns). This allows me to count the number different non-crawler IP addresses, grouped by CIK codes, which I use to estimate the average number of investors that follow one stock in a day, with the purpose of having a definition of a potential market. In particular in this research, I filter out crawlers and count the number of IPs grouped by stocks, using the EDGAR log files from 2015-01-01 to 2015-02-05.

It is important to note that, as I am comparing EDGAR users with Bloomberg Terminal users, who are humans, filtering out crawlers from EDGAR users is important. Therefore, I also use a filtered version of the data for my whole sample period (2010 - 2015), from The Software Repository for Accounting and Finance[2] of the University of Notre Dame, which provides an extension of the information used in Loughran and McDonald (2017), and counts with data from 2006 to 2015. This data have been filtered to eliminate damaged files[3], irrelevant entries, or those with missing CIK, accession number, IP, or date. Also, robot downloads (those with more than 49 downloads from a single IP within a single day or self identified as a web crawler), or with a server code larger or equal than 300, or records of traffic on the index page of a set of documents (e.g. index.htm) have been filtered out. The information set I use in this paper includes daily data, from 2010 to 2015, of the number of (exclusively) non-robot downloads for each stock, identified with the CIK number, with their respective dates. Filtering robot activity, is an advantage of using data on EDGAR searches, since this filtering is not possible to carry out when analyzing data related to traffic on other free sources of information, such that available at Google Trends or Wiki-media. Moreover, as the EDGAR database provides only hard financial information, the records of information acquisition at EDGAR are records more narrowed to investors' activities, when compared to, for example, Google Trends (Search Volume Index) which registers Google usage by potentially anybody (investors and non-investors).

As Bloomberg provides an Index of user activity, but not the number of users or number of searches, these data is not directly comparable to EDGAR downloads. I count the number of different stocks belonging to the CRSP, searched each day at each source of information and compare information acquisition activities between these two sources.

[1] https://www.sec.gov/dera/data/edgar-log-file-data-set.html
[2] https://sraf.nd.edu/
[3] All files from 2005-09-24 to 2006-05-10 were labeled by the SEC as "lost or damaged" (Loughran and McDonald, 2017).

a) Trends b) OLS Regression

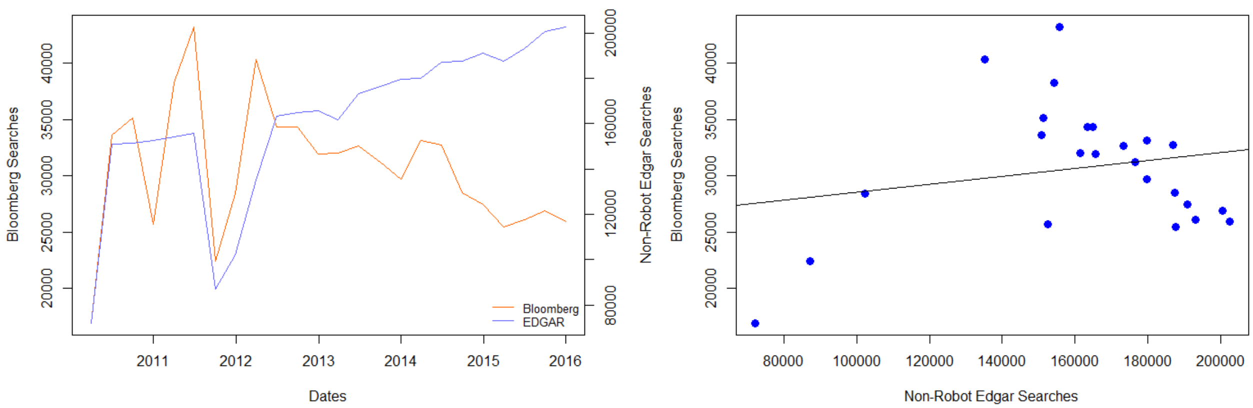

Figure 1: **Trends and Linear Regression of Information Acquisition Activities.** In panel a), the orange line represents, the quarterly sum of the daily number of stocks in the CRSP with a positive News Heat Index of Bloomberg, and the blue line represents the quarterly sum of the daily number of stocks in the CRSP with EDGAR searches. Panel b) shows the linear regression line between the two series of panel a).

Table 1: Summary Statistic on Search Activity at Each Source. Quarterly Number of Stocks in the CRSP Between 2010-03-31 and 2015-12-31.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
EDGAR	72,025	152,246	165,350	161,492	187,148	202,540
Bloomberg Terminals	16,952	26,679	31,575	30,684	33,811	43,181

In a first look at the data, from the trends of information acquisition activities at EDGAR and Bloomberg shown in panel a) of figure 1, we cannot get a clear idea of the substitution patterns. In the figure, the orange line represents, the quarterly sum of the daily number of stocks in the CRSP with a positive News Heat Index of Bloomberg, and the blue line represents the quarterly sum of the daily number of stocks in the CRSP with EDGAR searches. While the two series seem to follow opposite directions from around 2013 onward, they seem to move in a similar direction between 2010 and 2013, and this unclear pattern produces a non significant OLS slope estimate of 0.035 between the two series (panel b)).

I also collect data on service prices. EDGAR services are free of charge to the user and, as Bloomberg does not publicize its prices, I use a news media report containing historical subscription (two-year contract) prices per terminal from 2001 to 2013 of a single client in the US, which I project to 2015, using the average growth rate in price. I use an inflation-adjusted series, so that price changes reflect changes with respect to 2009 prices of Bloomberg subscriptions.

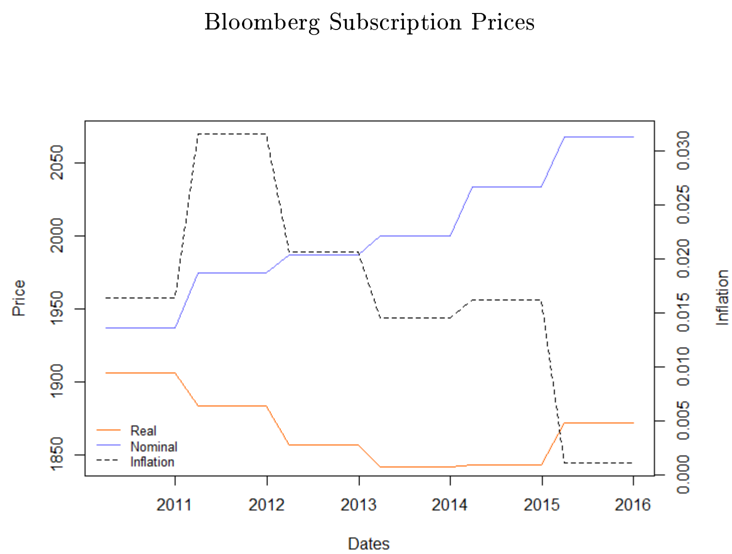

Figure 2: **Bloomberg prices 2010 - 2015.** The blue line represents two-year contract subscription prices per terminal; the orange line represents prices adjusted for inflation based on 2009 prices; the dashed line represents the annual inflation rate in the United States from fred.stlouisfed.org.

Interestingly, while the nominal service price has increased, the inflation-adjusted price of Bloomberg shows a slight decline. I assign a price of zero to EDGAR services and the inflation-adjusted price to Bloomberg services.

3.2 Demographics

I use the results of the National Financial Capability Study of 2012 and 2015 carried out by the Financial Industry Regulatory Authority (FINRA) Foundation in order to collect data on demographics. These are national self-report studies of the financial capability of American adults, which consists of a state-by-state online survey of 25,509 and 27,564 American adults in 2012 and 2015 respectively. In particular, I gather data on those individuals who responded "Yes" to the following question: "Not including retirement accounts, do you [does your household] have any investments in stocks, bonds, mutual funds, or other securities?" (variable $B14$[4]). From this population, I keep the answers on the variables of sex, ranges of annual income, willingness to take risks when investing, and the recorded answers to the following savings hypothetical problem: "Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow?" (variable $M6$). I use the answers to this question as a proxy for financial literacy, which provides an idea of the ability of investors to understand the implications of the information available to them.

[4] This is the only question related to investments in stocks.

Table 2: Summary Statistic on Observations for Demographic Draws.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
				2012		
"Male"	0.00	0.00	1.00	0.51	1.00	1.00
Income Levels	12,750	62,500	87,500	277,371	125,000	1'479,052
"Very Willing"	0.000	0.000	0.000	0.055	0.000	1.000
"More than $102"	0.00	1.00	1.00	0.85	1.00	1.00
				2015		
"Male"	0.00	0.00	1.00	0.54	1.00	1.00
Income Levels	12,750	62,500	87,500	253,809	125,000	1'479,052
"Very Willing"	0.000	0.000	0.000	0.081	0.000	1.000
"More than $102"	0.00	1.00	1.00	0.84	1.00	1.00

These variables are categorized with integers that are different from zero and one in many cases. For income, I transform the categories into levels of income (see table 2) using the average of the range for each category, where the wages of the the top 0.1% earners of the US population (2'808,104 USD) is the upper bound of the last category. With these income observations from the FINRA's survey, I calculate the logarithm of income levels, as well as the mean and standard deviation of the log-income for 2012 and 2015, parameters that I use to create a 100 draws of income from a log-normal distribution for each quarter, as shown in figure 3 and table 3. I denote the logarithm of income draws (which follows a normal distribution) as $Income$. For the other variables, I use dummies to obtain draws for my estimation, i.e. a dummy that takes the value of one for male, a dummy that equals one for "Very Willing", and a dummy that equals one for "More than $102." I use the average of each variable to generate 100 draws for each quarter from a binomial distribution, draws which I denote as $Male$, *Risk Lover* and *Literacy*.

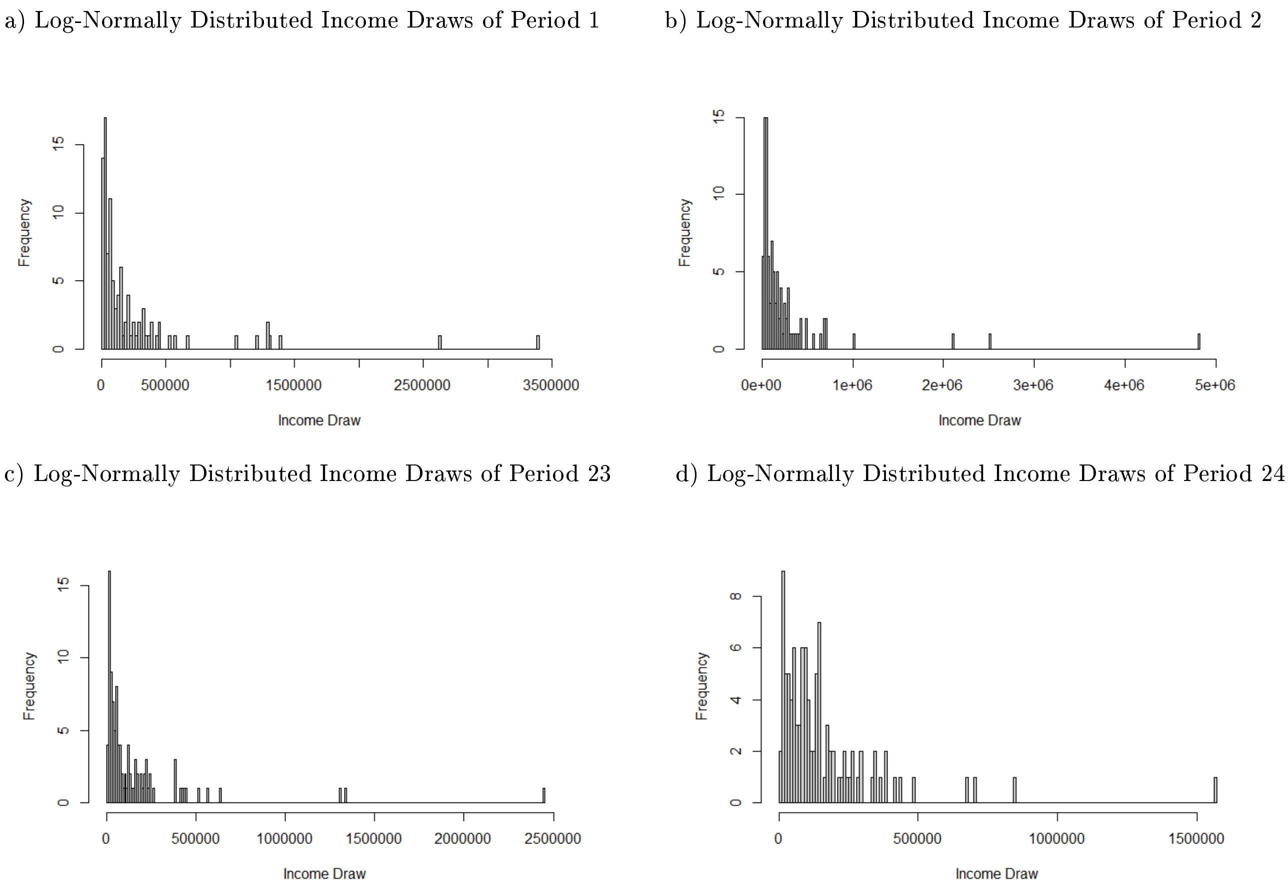

Figure 3: **Histograms of Log-Normally Distributed Income Draws.** Draws are obtained from a log-normal distribution, using the parameters estimated from survey observations.

Table 3: Summary Statistics on Log-Normally Distributed Income Draws of Four Quarters.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Period (market) 1	3,646	33,108	88,369	258,519	243,558	3'382,801
Period (market) 2	4,991	43,871	115,774	266,533	266,656	4'808,749
Period (market) 23	4,528	24,810	61,677	166,644	174,420	2'449,473
Period (market) 24	3,781	50,593	103,022	164,175	195,186	1'565,939

3.3 Data with Variation Across Alternatives

I calculate the cross-sectional average of stock returns, which I use to estimate the expected returns of the stocks sought through service j at quarter t, denoted $return_{jt}$. I interpret $return_{jt}$ as the expected attainable stock return from, or the expectations investors have on returns when, gathering financial information at provider j. Notice that the relationship between stock returns and information acquisition works in two directions: acquiring information on a set of stocks may influence their average returns, and having expectations on stock returns may have an effect on decisions of information acquisition.

I also calculate the cross-sectional average of the number of recommendations (in logs) issued by financial analysts, on stocks with activity at provider j, which I denote rec_{jt}. The number of stock recommendations in the literature is used as a proxy for analyst coverage (see e.g. Niehaus and Zhang, 2010). Furthermore, for each information provider, I calculate for each quarter, the average of the characteristics of the stocks sought at that source. I have in my sample, the data on firm size (log of market capitalization), return on assets (accounting variable), the volatility of earnings changes (accounting variable), stocks in the portfolios of institutional investors (list of section 13(f) securities of the SEC), and non-exhaustive sector dummies. I denote the average of these characteristics as $size_{jt}$, roa_{jt}, Δeps^{v}_{jt}, and D^{13F}_{ij}, and denote sectors with superscripts on the dummies such as D^{ID}_{jt} (industrials sector), D^{CG}_{jt} (consumer goods sector), D^{Tech}_{jt} (technology sector), D^{Fin}_{jt} (financials sectors), among others. In addition, my sample contains the projected stock prices from financial analysts from which I calculate the variance of forecast bias, i.e., of the difference between projected stock prices and realized stock prices, which captures the difficulty to evaluate the prospects of stocks (Hilary and Hsu's, 2013), and denote its average for the set of stocks sought at service j as $bvol_{jt}$.

Table 4: Summary Statistic on Alternative Specific Variables.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
$return_{jt}$	46	0.035	0.096	−0.222	−0.004	0.103	0.187
$F13_{ij}$	46	0.242	0.036	0.198	0.220	0.254	0.338
rec_{jt}	46	2.435	0.338	2.038	2.167	2.681	3.583
Δeps^{v}_{jt}	46	0.058	0.024	0.002	0.043	0.072	0.108
roa_{jt}	46	0.012	0.040	−0.063	−0.015	0.040	0.109
$size_{jt}$	46	8.167	0.889	6.966	7.487	8.851	11.396
$bvol_{jt}$	46	0.711	0.308	0.328	0.538	0.776	2.234
D^{CG}_{jt}	46	0.101	0.028	0.065	0.080	0.118	0.162
D^{CS}_{jt}	46	0.155	0.055	0.000	0.114	0.191	0.275
D^{MT}_{jt}	46	0.039	0.008	0.000	0.038	0.043	0.049
D^{Fin}_{jt}	46	0.175	0.055	0.056	0.131	0.221	0.266
D^{HC}_{jt}	46	0.133	0.035	0.000	0.116	0.154	0.188
D^{ID}_{jt}	46	0.170	0.022	0.111	0.156	0.186	0.204
D^{Tech}_{jt}	46	0.127	0.084	0.072	0.104	0.123	0.667
D^{TE}_{jt}	46	0.011	0.004	0.000	0.009	0.012	0.034
D^{UT}_{jt}	46	0.023	0.011	0.000	0.012	0.033	0.040

4 Empirical Strategy

I follow Berry, Levinsohn and Pakes (1995) and Nevo (2001). Two information providers offer their services in each of T quarters, and each market contains numerous investors (consumers of information) $i \in I$. In the estimation below, a market will be defined as a quarter. All investors within a given information market, face the same services with the same attributes so that the attributes of the information services may vary over markets but not over investors within each market. Let J_t be the number of options available to each investor in market $t \in T$. Variables of stock performance or characteristics related to service j in market t (attributes), such as the stock return attainable from seeking information at provider j in market t, are denoted by the K-dimensional row vector $\boldsymbol{x}_{jt}$; besides exogenous variables, this vector includes the endogenous stock return and service price. The unobserved attributes are denoted collectively as ξ_{jt} which, in this setup, represents the common utility that investors obtain from the unobserved attributes of service j in market t. I specify ξ_{jt} as the sum of two components: the mean valuation of the service characteristics that I (the researcher, as opposed to the investor) do not observe, ξ_j, and a quarter specific deviation from this mean, $\Delta\xi_{jt}$. I control for ξ_j, by including brand-specific dummy variables in the regressions. Market-specific components are included in $\Delta\xi_{jt}$ and are left as error terms. The utility that investor i in market t obtains from information provider j depends on observed and unobserved variables of stock performance (attributes) attainable from using the service. I assume that utility takes the form

$$U_{ijt} = \delta_{jt} + \boldsymbol{x}_{jt}\tilde{\boldsymbol{\beta}}_i + \varepsilon_{ijt} \tag{1}$$

where $\delta_{jt} = \boldsymbol{x}_{jt}\bar{\beta} + \xi_{jt}$, $\bar{\boldsymbol{\beta}} = (\bar{\beta}_1, ..., \bar{\beta}_k)'$ are parameters that are the same for all investors, $\tilde{\boldsymbol{\beta}}_i = (\tilde{\beta}_{i1}, ..., \tilde{\beta}_{ik})'$ are parameters that vary across investors, and ε_{ijt} is i.i.d. Type I extreme value with zero mean. Also,

$$\tilde{\boldsymbol{\beta}}_i = \varPi D_i + \varSigma \boldsymbol{v}_i \tag{2}$$

with $\boldsymbol{v}_i \sim N(0, I_K)$ (normal distribution of preferences over the K characteristics), where D_i is a $d \times 1$ vector of observed demographic variables, $\varPi$ is a $K \times d$ matrix of coefficients that measure how the taste characteristics vary with demographics, and $\varSigma$ is a scaling matrix so that $\mathbb{E}(v_{i,k}^2) = 1$. I assume investors observe all the service characteristics and take them into consideration when making decisions. Investors may decide not to purchase any of the services in my sample and, as usual, I normalize the mean utility of the outside option to zero.

4.1 Market Size and the Outside Option

In order to calculate market size, I use the aggregate number of stocks that belong to the CRSP, searched daily, during a quarter at each source of financial information. Subsequently, I multiply the aggregate number of stocks, by the average of daily counts of non-crawler IP addresses that access information on a stock (1,296). In order to obtain this estimate, I filter out crawlers and count the number of different IP addresses, grouped by CIK codes, that sought information at EDGAR using the 36 EDGAR log files in the period 2015-01-01 - 2015-02-05. Unfortunately, I do not have data on the daily number of Bloomberg terminal users who searched information at the stock or the aggregate level. Thus, I am assuming that those who access Bloomberg Terminals can also access EDGAR files, and then, this is a potential number for Bloomberg.

a) Securities Available at EDGAR b) Stocks Available at EDGAR in the CRSP

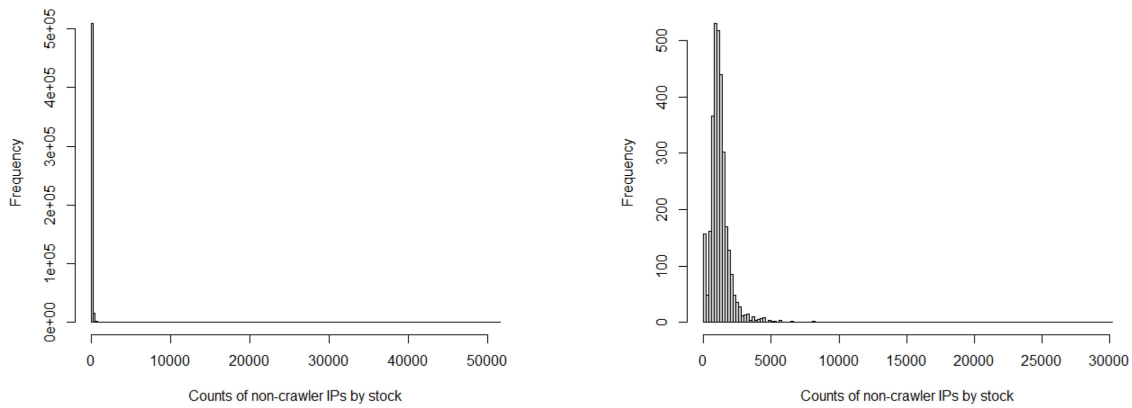

Figure 4: **Histograms of the number of different non-crawler IPs accessing data on a given stock, between 2015-01-01 and 2015-02-05.** In panel a), the data includes all securities available at EDGAR. Panel b) shows data on CRSP stocks only.

Table 5: Summary Statistic on Non-Crawler IP Counts by Stock, between 2015-01-01 and 2015-02-05.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Securities Available at EDGAR	1	4	10	41.73	22	51,406
Stocks Available at EDGAR in the CRSP	2	821	1,125	1,296	1,477	30,131

Counts of Non Crawler IPs Accessing EDGAR by Day

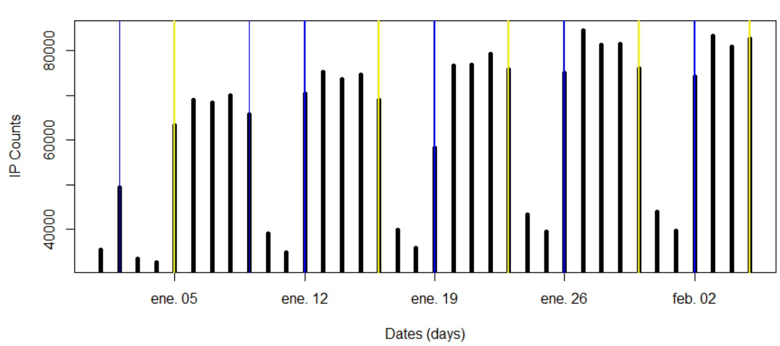

Figure 5: **Number of different non-crawler IPs accessing data at EDGAR by day, from 2015-01-01 to 2015-02-05.** Yellow lines represent Mondays and blue lines represent Fridays.

Around 32% of the US population have investments in stocks, bonds, mutual funds, or other securities outside of retirement accounts, percentage that has changed little over the years (34% in 2009, 32% in 2012, 30% in 2015 and 32% 2018), and from these, 74% own individual stocks in 2015 as well as in 2018 (see Lin et al., 2019a, b). There are sources of financial information alternative to internet-based sources such as radio and TV programs, printed news, interpersonal sources, workplace-based sources and personal financial advisors, and many investors do not acquire data often. For instance, Loibl and Hira (2009) analyze survey results of investors in the US, and identify that 30% of investors in their sample, are "reluctant" who reported to "seldom" or "never" gather financial information, whereas only 11% reported to obtain information "often" or "very often" from online sources. With 32% × 74% × 11% as the portion of the population (327'677,163) interested in online stock data, there are 8'535,334.7 online potential searches in a day, and 537'726,086 in a quarter. The definition of market shares I use is then the following,

$$Market\ Share_{jt} = \frac{Stocks_{jt} \times \bar{IPs}}{Potential\ Searches} \tag{3}$$

where $Stocks_{jt}$ is the aggregate daily number of stocks of the CRSP, searched at source j during quarter t, $\bar{IPs} = 1,296$ is the average count of non-crawler IP addresses that access information on a stock and $Potential\ Searches = 537'726,086$ is the number of potential searches in a quarter as defined before. The market share of the outside option is

$$1 - \left[\frac{Stocks_{jt} \times \bar{IPs}}{Potential\ Searches}\right] \tag{4}$$

Among the paid information providers in 2018, Bloomberg's market share was around 33%[5], Refinitiv Eikon's (formerly Reuters) was around 23%, Fact-Set's around 4% and Capital IQ's around 6%. Including other not observed sources of financial information, such as, radio and TV programs, printed news and other internet-based sources, it is sensible to observe that the outside option reaches a market share as high as 73,6% in the third quarter of 2011 (see table 6).

[5]See Wall Street Prep's article *Bloomberg vs. Capital IQ vs. FactSet vs. Thomson Reuters Eikon* and The Financial Times' article *Bloomberg and Reuters lose data share to smaller rivals*

Table 6: Market Shares In Number of Searches

Market shares are defined as the aggregate number of stocks that belong to the CRSP, searched in a quarter at each source of financial information, times the average count of non-crawler IP addresses that access information on a stock, divided by the population interested in online stock data.

Date	Product id	Shares	Date	Product id	Shares
2010-06-30	1	0.081	2013-06-28	1	0.079
2010-06-30	2	0.363	2013-06-28	2	0.418
2010-09-30	1	0.085	2013-09-30	1	0.075
2010-09-30	2	0.365	2013-09-30	2	0.426
2010-12-31	1	0.062	2013-12-31	1	0.072
2010-12-31	2	0.368	2013-12-31	2	0.433
2011-03-31	1	0.092	2014-03-31	1	0.080
2011-03-31	2	0.372	2014-03-31	2	0.433
2011-06-30	1	0.104	2014-06-30	1	0.079
2011-06-30	2	0.376	2014-06-30	2	0.451
2011-09-30	1	0.054	2014-09-30	1	0.069
2011-09-30	2	0.210	2014-09-30	2	0.452
2011-12-30	1	0.068	2014-12-31	1	0.066
2011-12-30	2	0.246	2014-12-31	2	0.460
2012-03-30	1	0.097	2015-03-31	1	0.061
2012-03-30	2	0.326	2015-03-31	2	0.452
2012-06-29	1	0.083	2015-06-30	1	0.063
2012-06-29	2	0.394	2015-06-30	2	0.466
2012-09-28	1	0.083	2015-09-30	1	0.065
2012-09-28	2	0.398	2015-09-30	2	0.484
2012-12-31	1	0.077	2015-12-31	1	0.062
2012-12-31	2	0.399	2015-12-31	2	0.488
2013-03-28	1	0.077			
2013-03-28	2	0.389			

4.2 Predicted Market Shares

Investors are assumed to use the service that gives the highest utility. This implicitly defines the set of unobserved variables that lead to the choice of information provider j:

$$A_{jt}(\boldsymbol{x}_t, \boldsymbol{\delta_t}, \Sigma, \Pi) = \{(D_i, \boldsymbol{v}_i, \boldsymbol{\varepsilon}_{it}) | U_{ijt} \geq U_{ilt} \quad \forall l = 0, 1, 2\}$$

where $\boldsymbol{\delta_t} = (\delta_{1t}, \delta_{2t})'$. Assuming ties occur with zero probability, the market share s_{jt} of the jth provider as a function of the mean utility levels of all the $J+1$ services, given the parameters, is

$$s_{jt}(\boldsymbol{x}_t, \boldsymbol{\delta_t}, \Sigma, \Pi) = \int_{A_{jt}} dP^*(\boldsymbol{\varepsilon}) dP^*(\boldsymbol{v}) dP^*(D) \tag{5}$$

where $P^*(\cdot)$ denotes population distribution functions. With the specified distributions for ε_{ijt} and $\boldsymbol{v_i}$, the market shares are

$$s_{jt} = \int \left[\frac{e^{(\boldsymbol{x}_{jt}\bar{\boldsymbol{\beta}} + \xi_j + \Delta\xi_{jt}) + (\boldsymbol{x}_j \tilde{\boldsymbol{\beta}}_i)}}{1 + \sum_j e^{(\boldsymbol{x}_{jt}\bar{\boldsymbol{\beta}} + \xi_j + \Delta\xi_{jt}) + (\boldsymbol{x}_j \tilde{\boldsymbol{\beta}}_i)}} \right] dP^*(\boldsymbol{v}) dP^*(D) \tag{6}$$

In my specification, from all the service characteristics that I include, namely $return_{jt}$, $price_{jt}$ and $F13_{jt}$, only the coefficient on $return_{jt}$ will be random. I estimate the parameters of the model by following the algorithm used by Berry, Levinsohn, and Pakes (1995), without a supply model as in Nevo (2001).

5 Identification

I control for the mean valuation of the service characteristics that I do not observe, ξ_j, by including a brand dummy $D_{jt}^{provider}$ that takes the value of one for Bloomberg. Furthermore, I address endogeneity problems of my regressors as follows. Service price affects information demand but prices are also a consequence of changes in demand, making this variable endogenous. I instrument subscription prices using its four-quarters lag as well as the first and second lags of Google's search volume index (SVI) on the expression "Bloomberg Terminal." Choi and Varian (2012) show that Google Trends queries help describing real variables such as travel destination planning and automobile sales. I argue that the popularity of Bloomberg Terminals is related to Bloomberg's capacity to charge higher prices for their services. Furthermore, I use Microsoft sales revenues as a proxy for Bloomberg service prices, since the Terminals are not available for MAC platforms and many functionalities involve an intensive use of Excel.

a) Google SVI on "Bloomberg Terminal" b) Microsoft Sales Revenues (Adjusted for Inflation)

Figure 6: **Service Price Instruments.** In panel a), I show the quarterly series of the lagged values of Google's SVI on the expression "Bloomberg Terminal." Panel b) shows the contemporaneous quarterly values of Microsoft sales adjusted for inflation.

In addition, as the demand for financial information may influence the expectations investors have on stock returns, I instrument $return_{jt}$ with a set of plausibly exogenous variables. In particular, I assume that decisions on information sources at a given quarter, do not affect the values of the average firm size

($size_{jt}$) or the average analyst coverage on stocks with activity at each provider (rec_{jt}), and I expect these variables to affect decisions on information acquisition only through their effects on expectations on stock returns.

I also include among instruments for $return_{jt}$, the mean of return on assets (roa_{jt}) which I argue, is an accounting variable not directly affected by decisions on information acquisition at a given quarter, and also the variance of earnings per share changes (Δeps^{v}_{jt}). Furthermore, I include the variance of the difference between projected stock prices and realized stock prices ($bvol_{jt}$), which captures the difficulty to evaluate the prospects of stocks. Among the instruments are also included, the average of (non-exhaustive) sector dummies of Consumer Goods (D^{CG}_{jt}), Consumer Services (D^{CS}_{jt}), Materials (D^{MT}_{jt}), Health Care (D^{HC}_{jt}), Industrials (D^{ID}_{jt}), Technology (D^{Tech}_{jt}), Financials (D^{Fin}_{jt}), Telecomunications (D^{TE}_{jt}) and Utilities (D^{UT}_{jt}) which are exogenous. I argue that these variables affect the choice of information provider, only through their effect on stock return expectations. Additionally to the contemporaneous variables, I include the first lag of roa_{jt}, Δeps^{v}_{jt}, $bvol_{jt}$, rec_{jt} and $size_{jt}$.

Table 7: Linear Correlation With Expected Returns

	Estimate	p-value
$return_{jt}$	1	0
rec_{jt}	0.117	0.44
$rec_{j,t-1}$	0.024	0.87
$size_{jt}$	0.181	0.23
$size_{j,t-1}$	-0.056	0.71
roa_{jt}	0.115	0.44
$roa_{j,t-1}$	0.091	0.55
Δeps^{v}_{jt}	-0.099	0.51
$\Delta eps^{v}_{j,t-1}$	-0.005	0.97
$bvol_{jt}$	-0.023	0.88
$bvol_{j,t-1}$	0.173	0.25

With the purpose of analyzing the instruments, I estimate the conditional first-stage F statistic of Angrist and Pischke (2009). I regress service prices on instruments, and in turn regress the fitted values, on exogenous variables (brand and 13(f) list dummies) and on the fitted values, of the regression of expected returns on the set of instruments. Subsequently, I regress the last residuals on instruments and estimate the F-statistic to have an idea of the goodness of instruments for price. In order to analyze the instruments for the return expectations, I carry out a similar exercise in which the first regression contains expected returns as the dependent variable. The results of the estimations of the first regressions on instruments are in table 8. As a rule of thumb, values larger than 10 of Angrist and Pischke's (2009) F-statistic, are a good indicator that instruments are not weak, which is the case for the regressions of service prices and expected returns.

Table 8: Linear Regression of Residuals on Instruments.
Variables svi_{t-1} and $msft_t$ are Google's SVI of the expression "Bloomberg Terminal" and Microsoft Sales respectively. Prices and sales adjusted for annual inflation starting from 2010. Note: *p<0.1; **p<0.05; ***p<0.01.

	Dependent variable:	
	$service\ price_{jt}$	$return_{jt}$
	(1)	(2)
$price_{jt-4}$	0.994***	0.0001
	(0.016)	(0.0001)
svi_{jt-1}	−0.397	0.001
	(0.423)	(0.003)
svi_{jt-2}	−0.899**	−0.003
	(0.365)	(0.002)
$msft_{jt}$	1.693	0.005
	(1.219)	(0.008)
roa_{jt}	190.573	−0.622
	(272.451)	(1.734)
roa_{jt-1}	−30.621	−0.291
	(206.526)	(1.314)
$bvol_{jt}$	24.660	−0.138
	(16.908)	(0.108)
$bvol_{jt-1}$	0.984	0.020
	(7.233)	(0.046)
$size_{jt}$	20.503	0.605***
	(15.644)	(0.100)
$size_{jt-1}$	13.848	−0.502***
	(14.635)	(0.093)
rec_{jt}	19.431	−1.395***
	(49.741)	(0.317)
rec_{jt-1}	−46.109	1.443***
	(47.224)	(0.301)
Δeps^v_{jt}	−599.608***	−1.289
	(193.182)	(1.229)
Δeps^v_{jt-1}	20.453	0.561
	(221.537)	(1.410)
D^{CG}_{jt}	−816.908	0.179
	(493.782)	(3.142)
D^{CS}_{jt}	−794.252*	2.016
	(383.213)	(2.439)
D^{MT}_{jt}	157.202	3.522
	(511.184)	(3.253)
D^{Fin}_{jt}	−282.980	−0.755
	(336.189)	(2.140)
D^{HC}_{jt}	−446.065	2.869
	(402.882)	(2.564)
D^{ID}_{jt}	−662.113**	2.140
	(308.803)	(1.965)
D^{Tech}_{jt}	−693.811**	1.267
	(302.983)	(1.928)
D^{TE}_{jt}	−2,270.632*	5.802
	(1,118.614)	(7.119)
D^{UT}_{jt}	1.689	0.038
	(628.467)	(4.000)
Constant	335.217	−2.227
	(316.974)	(2.017)
Observations	46	46
Residual Std. Error (df = 22)	8.990	0.057
F Statistic (df = 23; 22)	21,539.810	4.529
Angrist-Pischke F Statistic	12.770	130.014

6 Results

I estimate a full random coefficients model in order to control for unobserved heterogeneity, using several specifications. The results are in table 9. The statistically negative estimate on $price_{jt}$ suggests that lower prices induce a higher demand for financial information services, as one would expect from the standard economic intuition. Moreover, the statistically positive estimate on $return_{jt}$, suggests that higher expectations on the stock returns attainable when searching at a source of information, induce a higher demand for the services of this provider. The statistically negative estimate on $A8_i \times return_{jt}$ suggest that the marginal valuation of expected returns decreases with higher income levels.

Table 9: Results: Full Random Coefficients Model.

Instruments for $return_{jt}$ are roa_{jt}, Δeps^v_{jt}, $bvol_{jt}$, rec_{jt}, $size_{jt}$ and their one-quarter lags, as well as $D^{CG}_{jt}, D^{CS}_{jt}, D^{MT}_{jt}, D^{HC}_{jt}, D^{ID}_{jt}, D^{Tech}_{jt}, D^{Fin}_{jt}, D^{TE}_{jt}$ and D^{UT}_{jt}. Bloomberg service price is instrumented with the one-quarter lag and the two-quarter lag of Google's SVI of the expression "Bloomberg Terminal", as well as with a four-quarters lag of price and the contemporaneous value of Microsoft's sales, both adjusted for annual inflation starting from 2010. Codes in parenthesis are the labels in the survey, used to obtain the corresponding draws. **Computational Details:** market shares are integrated with modified latin hypercube sampling draws (MLHS) and 100 demographic draws; method for standard errors: heteroskedastic.

	(1)	(2)	(3)	(4)
(Intercept)	1.90***	1.900***	2.000***	1.900***
	(0.520)	(0.610)	(0.600)	(0.570)
$return_{jt}$	103.00*	103.000*	102.000*	103.000*
	(54.000)	(55.000)	(54.000)	(59.000)
$price_{jt}$	-0.008**	-0.008**	-0.008**	-0.008**
	(0.004)	(0.004)	(0.004)	(0.004)
$F13_{jt}$	-8.50***	-8.400***	-8.600***	-8.200***
	(2.100)	(2.600)	(2.600)	(2.400)
$D^{provider}_{jt}$	13.00*	13.000*	13.000*	13.000*
	(6.600)	(6.700)	(6.900)	(6.600)
Demographics (Random Coefficients)				
(Unobserved) $\sigma \times return_{jt}$	15.0**	15.0**	15.0**	15.0*
	(7.200)	(7.500)	(7.200)	(7.900)
(A8) $Income_i \times return_{jt}$	-9.0*	-9.0*	-9.0*	-9.0*
	(4.700)	(5.200)	(5.300)	(5.300)
(M6) $Literacy_i \times return_{jt}$	–	-0.400	-0.400	-0.400
	–	(20.000)	(20.000)	(17.000)
(A3) $Male_i \times return_{jt}$	–	–	2.0	2.0
	–	–	(10.000)	(12.000)
(J2) $Risklover_i \times return_{jt}$	–	–	–	-13.0
	–	–	–	(39.000)

6.1 Robustness Checks

To calculate market sizes, in the previous section I used the average, of daily counts of non-crawler IP addresses that access information on a stock, with information from 36 EDGAR log files, as well as the number of stocks searched each period. Yet, the average number of times that unique visitors, access the system during a quarter, per stock, could be not the same for EDGAR and Bloomberg. 12.7% of respondents of the Investor Survey of 2015[6], which is a separate follow-up survey of investors from the 2015

[6] Available only for 2015 and 2018.

NFCS, reported to have used paid online services during the past 12 months (13.38% in 2018 reported to have used paid subscription services). With a market share of 33% among subscription providers, and 537'726,086 online potential searches in a quarter, Bloomberg's market size is 22'536,100.26. Meanwhile, the number of non-robot searches at EDGAR was 18'462,680 in the fourth quarter of 2015, for which, in terms of visits, Bloomberg market size is 1.22 times the EDGAR size. In this subsection, I show the results of multiplying Bloomberg's unique visitors per stock, by 1.22 (see market shares in table 11 of the appendix). The estimates shown in table 10, remain statistically positive for $return_{jt}$ and statistically negative for $price_{jt}$. Moreover, the estimates on the interactions between income and return are significant in all specifications.

Table 10: Results: Full Random Coefficients Model.

Instruments for $return_{jt}$ are roa_{jt}, Δeps^{v}_{jt}, $bvol_{jt}$, rec_{jt}, $size_{jt}$ and their one-quarter lags, as well as $D^{CG}_{jt}, D^{CS}_{jt}, D^{MT}_{jt}, D^{HC}_{jt}, D^{ID}_{jt}, D^{Tech}_{jt}, D^{Fin}_{jt}, D^{TE}_{jt}$ and D^{UT}_{jt}. Bloomberg service price is instrumented with the one-quarter lag and the two-quarter lag of Google's SVI of the expression "Bloomberg Terminal", as well as with the contemporaneous value of Microsoft's sales, both adjusted for annual inflation starting from 2010. Codes in parenthesis are the labels in the survey, used to obtain the corresponding draws. **Computational Details:** market shares are integrated with modified latin hypercube sampling draws (MLHS) and 100 demographic draws; method for standard errors: heteroskedastic.

	(1)	(2)	(3)	(4)
(Intercept)	1.303***	1.216***	1.318***	1.377***
	(0.301)	(0.311)	(0.302)	(0.325)
$return_{jt}$	102.446**	104.618**	102.021*	102.417*
	(44.960)	(47.700)	(61.520)	(53.320)
$price_{jt}$	-0.005*	-0.006**	-0.005*	-0.006**
	(0.003)	(0.003)	(0.003)	(0.003)
$price_{jt-4}$	0.002	0.004*	0.002)	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
$F13_{jt}$	-5.549***	-5.125***	-5.636***	-5.797***
	(1.246)	(1.289)	(1.282)	(1.283)
$D^{provider}_{jt}$	3.100	3.223	3.614	4.705
	(4.368)	(4.324)	(4.599)	(4.710)
Demographics (Random Coefficients)				
(Unobserved) $\sigma \times return_{jt}$	5	5	5	5
	(11.589)	(11.613)	(13.415)	(13.689)
(A8) $Income_i \times return_{jt}$	-9**	-9**	-9*	-9*
	(3.952)	(4.546)	(5.287)	(4.663)
(M6) $Literacy_i \times return_{jt}$	–	-3	-3	-3
	–	(11.409)	(14.254)	(15.435)
(A3) $Male_i \times return_{jt}$	–	–	6	6
	–	–	(17.301)	(16.245)
(J2) $Risk\ lover_i \times return_{jt}$	–	–	–	-13
	–	–	–	(17.773)

7 Discussion of Results and Limitations

These results were obtained from data on the daily number of stocks, sought by investors through each source of information, and then transformed in number of investors by using the number of IP addresses per stock in EDGAR. My data does not capture all differences of information acquisition between sources, since up to now I have not included the number of user accounts of Bloomberg terminals in my sample. Ideally, the variation in information acquisition between providers in the data would be a consequence of observing the number of users with activity at each source of information, and the type of security that was of each user's interest, so that I could associate investor choice of information provider directly to his portfolio characteristics.

In order to capture differences of information acquisition between sources in a more direct manner, I will look for data on Bloomberg users. Bloomberg counts with identity objects that contain information on, e.g. the last time the user logged in or the functions users are accessing, information that is used by Bloomberg's sales force through the function UUID (not available for final users). Incorporating this data in my sample will allow me to compare search activity at each source in a direct manner and to capture all differences in information acquisition. Furthermore, I will estimate market shares from media attention, as information providers that receive more attention should count with higher search activity. In particular, I will count the number of times Bloomberg and EDGAR appear in the crowd-sourced content service Seeking Alpha, which has over 16,000 contributors who share finance news and investing ideas, covering around 8,000 tickers in an archive of over 1 million articles as of 2021. Web-scrapping *seekingalpha.com* and counting the number of times contributors make a reference to a source of information, has important advantages over downloading data from the more traditional source of media attention Google Trends. First, a great amount of Google's records for searches on the words "EDGAR" or "Bloomberg" is not related to financial information acquisition as many Google users are not investors interested in the prospect of firms, but are non investors curious about these platforms and their workings. Second, the number of references per period of time is more intuitive than the Search Volume Index which takes values from 0 to 100, where a value of 50 is vaguely described as "half of popularity" by Google.

Also, my data on the costs of using each provider, included historical information on only one Bloomberg client, but, besides the fact that there are several clients around the world, the cost of acquiring data varies with the price of technology even for open access sources. Including data on more clients of Bloomberg as well as data on technology, will offer me better inputs to estimate the effects of prices on financial information acquisition.

Two important counterfactual questions can be answered with this research. The first one, related to the fact that open access sources of information also provide hard financial data (besides news), consists in understanding whether the market will eventually consider information on accounting variables, provided by subscription services, as completely substitutable, or in question form: would searches of financial statements at Bloomberg be completely absorbed by EDAGAR in the event that Bloomberg disappears from the market? Would be there investors who stop being market participants if the subscription provider with the highest market share disappears? The second one, is related to the fact that many investors, use EDGAR not only because it contains relevant financial information but also because it does not require a subscription. What would happen to search activities if EDGAR became a subscription provider and its access price increases?

8 Conclusions

The widespread access to internet, offers to investors a large set of choice alternatives of financial information providers, which comprehends subscription services as well as no-fee open access providers. In this paper, I present empirical evidence that investors substitute information sources according to subscription prices, as well as to stock return expectations and investor's income.

Understanding investors' choices over sources is important for information vendors as well as for investors' welfare. Standard economic theory tells us that stock prices and stock returns convey information from the informed investors to the uninformed, in a degree that depends on the number of individuals who are informed. While having greater amounts of information delivered through many providers improves the quantity and diversity of information available to investors, it comes with its new own difficulties for decision making: cognitive resources and attention are scarce resources that must be allocated. This research contributes to the literature on information acquisition in financial markets, presenting empirical evidence that help to understand decisions over financial data vendors.

Appendix

Table 11: Market Shares for Robustness Check

EDGAR market size is defined as the aggregate of the daily number of stocks that belong to the CRSP, searched in a quarter, times the average count of non-crawler IP addresses that access information on a stock, For Bloomberg, the average count of IP addresses is estimated using that of EDGAR, times 1.22. Market shares equal market sizes divided by the population interested in online stock data.

Date	Product id	Shares	Date	Product id	Shares
2010-06-30	1	0.099	2013-06-28	1	0.096
2010-06-30	2	0.363	2013-06-28	2	0.418
2010-09-30	1	0.103	2013-09-30	1	0.092
2010-09-30	2	0.365	2013-09-30	2	0.426
2010-12-31	1	0.076	2013-12-31	1	0.087
2010-12-31	2	0.368	2013-12-31	2	0.433
2011-03-31	1	0.113	2014-03-31	1	0.097
2011-03-31	2	0.372	2014-03-31	2	0.433
2011-06-30	1	0.127	2014-06-30	1	0.096
2011-06-30	2	0.376	2014-06-30	2	0.451
2011-09-30	1	0.066	2014-09-30	1	0.084
2011-09-30	2	0.210	2014-09-30	2	0.452
2011-12-30	1	0.084	2014-12-31	1	0.081
2011-12-30	2	0.246	2014-12-31	2	0.460
2012-03-30	1	0.119	2015-03-31	1	0.075
2012-03-30	2	0.326	2015-03-31	2	0.452
2012-06-29	1	0.101	2015-06-30	1	0.077
2012-06-29	2	0.394	2015-06-30	2	0.466
2012-09-28	1	0.101	2015-09-30	1	0.079
2012-09-28	2	0.398	2015-09-30	2	0.484
2012-12-31	1	0.094	2015-12-31	1	0.076
2012-12-31	2	0.399	2015-12-31	2	0.488
2013-03-28	1	0.094			
2013-03-28	2	0.389			

References

Angrist, J., and Pischke, J.S., (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton.

Asthana, S., Balsam, S., and Sankaraguruswamy, S. (2004). Differential response of small versus large investors to 10-K filings on EDGAR. *The Accounting Review*, 79(3), 571-589.

Baum, J. A., Bowers, A., and Mohanram, P. (2016). Mutual forbearance and competition among security analysts. *Management Science*, 62(6), 1610-1631.

Behrendt, S., Peter, F. J., and Zimmermann, D. J. (2020). An encyclopedia for stock markets? Wikipedia searches and stock returns. *International Review of Financial Analysis*, 101563.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841-890.

Chi, S. and Shanthikumar, D. M. (2018). Do Retail Investors Use SEC Filings? Evidence from EDGAR Search (October 25, 2018). Available at SSRN

Choi, H., and Varian, H. (2012). Predicting the present with Google Trends. *Economic record*, 88, 2-9.

Coeurdacier, N., and Rey, H. (2013). Home bias in open economy financial macroeconomics. *Journal of Economic Literature*, 51(1), 63-115.

Da, Z., Engelberg, J., and Gao, P. (2011). In search of attention. *The Journal of Finance*, 66(5), 1461-1499.

Gabaix, X., Laibson, D., Moloche, G., and Weinberg, S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review*, 96(4), 1043-1068.

Grossman, S. J., and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, 70(3), 393-408.

Hilary, G., and Hsu, C. (2013). Analyst forecast consistency. *the Journal of Finance*, 68(1), 271-297.

Hu, Y., Li, X., Goodell, J. W., and Shen, D. (2021). Investor attention shocks and stock co-movement: substitution or reinforcement?. *International Review of Financial Analysis*, 73, 101617.

LaValle, I. H. (1968). On cash equivalents and information evaluation in decisions under uncertainty Part I: Basic theory. *Journal of the American Statistical Association*, 63(321), 252-276.

Loughran, T. and McDonald, B. (2017). The use of EDGAR filings by investors. *Journal of Behavioral Finance*, 18(2), 231-248.

Loibl, C., and Hira, T. K. (2009). Investor information search. *Journal of economic psychology*, 30(1), 24-41.

Lipman, B. L. (1991). How to decide how to decide how to...: Modeling limited rationality. *Econometrica: Journal of the Econometric Society*, 1105-1125.

Lin, J., Bumcrot, C., Ulicny, T., Mottola, G., Walsh, G., Ganem, R., Kieffer, C. and Lusardi, A. (2019a). The State of US Financial Capability: The 2018 National Financial Capability Study. FINRA Investor Education Foundation.

Lin, J. T., Bumcrot, C., Mottola, G., and Walsh, G. (2019b). Investors in the United States: A report of national financial capability study. FINRA: Investor Education Foundation.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, 42(3), 483-510.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2), 307-342.

Niehaus, G., and Zhang, D. (2010). The impact of sell-side analyst research coverage on an affiliated broker's market share of trading volume. *Journal of Banking & Finance*, 34(4), 776-787.

Peng, L. and Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, 80(3), 563-602.

Sadler, E. (2021). A Practical Guide to Updating Beliefs From Contradictory Evidence. *Econometrica*, 89(1), 415-436.

Schreiber, T. (2000). Measuring information transfer. *Physical review letters*, 85(2), 461.

Stock, J. H., Wright, J. H., and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20(4), 518-529.

Vlastakis, N., and Markellos, R. N. (2012). Information demand and stock market volatility. *Journal of Banking & Finance*, 36(6), 1808-1821.

Wang, Pingle. (2020). Demand for Information and Stock Returns: Evidence from EDGAR. Available at SSRN

Willinger, M. (1989). Risk aversion and the value of information. *Journal of Risk and Insurance*, 320-328.