University of Toronto
Department of Economics

Working Paper 326

The Determinants of International Investment and Attention Allocation: Using Internet Search Query Data

By Jordi Mondria, Thomas Wu, and Yi Zhang

June 11, 2007

The Determinants of International Investment and Attention Allocation: Using Internet Search Query Data*

Jordi Mondria† Thomas Wu‡ Yi Zhang§

University of Toronto UC Santa Cruz UC Santa Cruz

August 2008

Abstract

The challenge of evaluating asymmetric information theories relies on assessing the pieces of information investors decide to process. This paper overcomes such challenge exploring a unique dataset containing the "search/click-through" behavior of internet search engine users. We analyze the relationship between attention allocation and international investment decisions by combining U.S. data on portfolio holdings of foreign securities with the attention allocated by America Online customers in search queries towards these countries. We find evidence that: (i) agents tend to search more information about countries where they hold more assets, and (ii) agents tend to invest more in countries where they process more information.

Keywords: Foreign Asset Holdings, Attention Allocation , Internet Search Query.

JEL Codes: F30, D82, G11.

*The authors would like to thank comments and suggestions from Ricardo Reis, and seminar participants at UC Davis, Getulio Vargas Foundation - Sao Paulo and the National Taiwan University.

†University of Toronto, Department of Economics. E-mail: jordi.mondria@utoronto.ca.

‡UC Santa Cruz, Department of Economics. E-mail: thomaswu@ucsc.edu.

§UC Santa Cruz, School of Engineering. E-mail: yiz@soe.ucsc.edu.

1 Introduction

Recent studies have rationalized many puzzles in finance and economics by allowing agents to choose what to learn. The empirical challenge of this literature relies on the difficulty to measure the pieces of information each agent processes. This paper overcomes such challenge exploring a unique dataset containing the "search/click-through" behavior of internet search engine users. We analyze the relationship between attention allocation and international investment decisions by combining U.S. data on portfolio holdings of foreign securities with the attention allocated by 657,426 America Online (henceforth AOL) customers in search queries towards these countries.

The idea that "*most people everyday encounter, or could very easily encounter, much more information that is in principle relevant to their economic behavior than they actually respond to*" (Sims, 2006) - known as inattentiveness - has recently been incorporated into finance and economic models. The basic implication of this assumption is that agents have to decide the relevant information for their decision-making. In macroeconomics, inattentiveness helps explaining sticky prices in Mackowiak and Wiederholt (2007) and Mankiw and Reis (2002) and consumption dynamics in Luo (2007). Falkinger (2007) develops a theoretical model in which firms have to compete first for consumers' attention before competing for their budgets. In finance, it has been used to better understand the equity-premium puzzle in Gabaix and Laibson (2002), contagion across emerging economies in Mondria (2007), portfolio under-diversification in Van Nieuwerburgh and Veldkamp (2008a), home equity bias in Van Nieuwerburgh and Veldkamp (2008b) and the forward discount bias in Bacchetta and Van Wincoop (2007).

Due to limitations on datasets, the supporting evidence of these models is questionable. Several models rely on calibration and simulation. One major criticism to this technique is with respect to the parameters that refer to the limits in the information processing capacity. Other papers rely on regressions with variables that proxy attention allocation. As a measure of information flow, Portes and Rey (2005) uses volume of call traffic and Mondria and Quintana-Domeque (2006) uses the number of stories about a country in the newspaper. On a

third different approach, Gabaix et al (2003) studies attention allocation on an experimental lab. However, the degree of noise that separates these proxy variables to the unobservable information choice decisions may generate some skepticism towards the empirical results.

Measuring attention allocation is a complicated task. Agents process information using different media (i.e., books, magazines, newspapers, radio, TV, the internet). For most of these different type of media, it is virtually impossible to track the amount of effort spent by the agent on processing information about different topics. On August 3, 2006, AOL released a dataset that includes over 21 million web search queries from 657,426 customers. A web search query is the exact phrase that a user types into the search engine to satisfy his or her information needs. The data includes all searches from those users for a three month period (March 1st to May 31st of 2006), as well as whether they clicked on a result and where it appeared on the result page. In this paper, we explore the AOL dataset to help us measure the attention allocated to different countries. We believe that a measure based on the internet search queries can proxy attention allocation because: first, the World Wide Web is becoming the predominant information media; second, search engines are the most popular tool to help users find reliable information on the "Web" since it minimizes the time required and the amount of information which must be consulted; third, by having the exact search query, we know what is the topic that the user was interested in.

Using this dataset, we constructed a click-through series that counts the number of times a user clicked through a search result from a particular country. The idea is to measure the attention allocated to a country by the number of times this country provided the answer to a search query. We combine the click-through series with data from the 2006 survey of U.S. portfolio holdings of foreign securities, sponsored by the Department of the Treasury, the Federal Reserve Bank of New York, and the Board of Governors of the Federal Reserve System. The survey measured U.S. holdings of foreign securities at year-end 2006 of approximately $6.0 trillion. Figure 1 reveals a very close relationship between investment and attention allocation on foreign equities.

[Insert Figure 1 about here]

The relationship between U.S. foreign equity holdings as of 31 of December of 2006 and the AOL country attention allocation between March and May of 2006 can be potentially explained by three hypotheses:

i. No causality, just correlation caused by the development level in each country;

ii. Causality runs from asset holding to attention: other things equal, agents tend to allocate more attention to the countries where they hold more assets;

iii. Causality runs from attention to holdings: other things equal, agents tend to invest more in countries from which they process more information.

While the first explanation is the most obvious one, the other two are by far more interesting. In this paper, we build a model with inattentive investors that rationalizes the two-way causal relationship between attention allocation and international investment. We find empirical evidence that the strong relationship between asset holdings and attention allocation is actually a result of all three factors. First, there is correlation caused by a country's development level. This effect is captured by the positive effect of GDP per capita on both the attention allocation and also the amount of securities held in each country. Agents not only invest more but also allocate more attention to more developed economies. Second, causality runs from asset holdings to attention allocation. Using instrumental variables that capture the level of development in financial markets and economic proximity, we find that agents do allocate more attention to the countries where they hold more assets. Third, causality also runs from attention allocation to asset holdings. Using instrumental variables that are related to a country's popularity among internet users, we find that international investors tend to hold more assets at more familiar countries. The rest of the paper is organized as follows. Section 2 develops the model. Section 3 describes the dataset. Section 4 shows the empirical results. Section 5 concludes.

2 Model

The objective of the model is to show the interaction between asset holdings and attention allocation. In particular, we are interested in finding the effect of attention allocation on asset

holdings and the reverse, the effect of asset holdings on attention allocation.

2.1 Model Description

This is a partial equilibrium model with a continuum of investors and two countries. The economy consists of a risk free asset, which pays R units of the consumption good, and two independent risky assets. Even though this is a static model, four discrete events occur during the operation of the market. First, traders are endowed with an initial wealth W_{i0} and limited information processing resources κ. Second, investors allocate their limited information processing resources to analyze both stock markets. Third, each investor decides the optimal asset holdings $X_i = (x_{i,1}, x_{i,2})'$ given private information about each market, $\tilde{Y}_i = (\tilde{y}_{i,1}, \tilde{y}_{i,2})'$, which depends on the amount of information processed about each stock market. Fourth, trading ceases and investors consume the payoff realized from their portfolio.

Investors, with absolute risk tolerance parameter ρ, maximize a mean-variance objective function

$$U_i = E\left(E\left[W_i' \mid \tilde{Y}_i\right] - \frac{1}{2\rho} Var\left[W_i' \mid \tilde{Y}_i\right]\right)$$

subject to the following budget constraint

$$W_{i1} = W_{i0} + X_i'(\tilde{R} - RP)$$

where W_{i0} is the initial wealth of agent i, $X_i = (x_{i,1}, x_{i,2})'$ is the asset holdings vector of agent i, $\tilde{R}$ is the vector of risky asset payoffs, $\tilde{Y}_i = (\tilde{y}_{i,1}, \tilde{y}_{i,2})'$ is agent's i private information about each market and $\tilde{P}$ is the price vector of the risky assets.

Investors devote information capacity to process information about the vector of unknown and independent asset payoffs $\tilde{R} = (\tilde{r}_1, \tilde{r}_2)' \sim N(0, \Sigma_R)$. Trader i receives a private signal about each risky asset $j = 1, 2$ given by

$$\tilde{y}_{i,j} = \tilde{r}_j + \tilde{\varepsilon}_{i,j} \text{ where } \varepsilon_{i,j} \sim N(0, \sigma_{i,j}^2)$$

Investors want to obtain information about the risky assets in order to reduce the uncertainty

of their optimal portfolio. They face the following linear technology on processing information about the state of the economy

$$\frac{1}{\sigma_{i,j}^2} = \frac{\kappa_{i,j}}{\sigma_{r,j}^2}$$

The precision of a private signal is higher if more attention is allocated to that particular signal. However, investors have a limited capacity to process information about asset payoffs.

$$\kappa_{i,1} + \kappa_{i,2} \leq \kappa \text{ and } \kappa_{i,j} \geq 0 \text{ for } j = 1, 2$$

where $\kappa_{i,1}$ and $\kappa_{i,2}$ are the attention allocated to each stock market. The information constraint imposes a limit in the reduction of the agent i's uncertainty about the future payoff of the risky asset j. The information processing resources have to be optimally divided between the two risky assets. After deciding the amount of information to be processed about each stock market, investors incorporate the information from the private signal, $\tilde{Y}_i$, into their beliefs through Bayesian updating. Then, investors derive their posterior beliefs about the asset payoffs and decide their optimal asset holdings.

2.2 Effect of Attention Allocation on Asset Holdings

The objective of this section is to show the effect of attention allocation on asset holdings. Suppose each agent chooses the optimal risky asset demand taking as given any attention allocation. After observing a given private signal about each risky asset, investors derive their posterior beliefs about the asset payoffs in order to choose their optimal asset holdings

$$X_i\left(\tilde{Y}_i\right) = \rho Var\left[\tilde{R} \mid \tilde{Y}_i\right]^{-1} E\left[\tilde{R} - RP \mid \tilde{Y}_i\right]$$

Asset holdings are increasing with attention allocation, $\frac{\partial x_{i,j}}{\partial \kappa_{i,j}} \geq 0$ when expected excess returns are positive since

$$\frac{\partial Var_{jj}\left[\tilde{R}|\tilde{Y}_i\right]^{-1}}{\partial \kappa_{i,j}} \geq 0 \quad \text{where} \quad Var\left[\tilde{R} \mid \tilde{Y}_i\right] = \begin{pmatrix} \frac{\sigma_{r1}^2}{1+\kappa_{i,1}} & 0 \\ 0 & \frac{\sigma_{r2}^2}{1+\kappa_{i,2}} \end{pmatrix}$$

Therefore, investors hold more assets in countries where they allocate more attention.

2.3 Effect of Asset Holdings on Attention Allocation

The objective of this section is to show the effect of asset holdings on attention allocation. Suppose that investors take as given their asset holdings when choosing where to allocate their information resources, then the objective function becomes

$$\min_{\kappa_{i,1},\kappa_{i,2}} X_i' V\left(\tilde{R} \mid \tilde{Y}_i\right) X_i$$

which can be rewritten as

$$\min x_{i1}^2 \left(\frac{\sigma_{r1}^2}{1+\kappa_{i,1}}\right) + x_{i2}^2 \left(\frac{\sigma_{r2}^2}{1+\kappa_{i,2}}\right)$$

subject to the information processing constraint

$$\kappa_{i,1} + \kappa_{i,2} \leq \kappa \text{ and } \kappa_{i,j} \geq 0$$

The optimal attention allocation is given by

$$\kappa_{i,1} = \begin{cases} \kappa & \text{if } x_{i1}\sigma_{r1} \geq (1+\kappa)\, x_{i2}\sigma_{r2} \\ 0 & \text{if } x_{i2}\sigma_{r2} \geq (1+\kappa)\, x_{i1}\sigma_{r1} \\ \frac{x_{i1}\sigma_{r1}(1+\kappa) - x_{i2}\sigma_{r2}}{x_{i1}\sigma_{r1} + x_{i2}\sigma_{r2}} & \text{otherwise} \end{cases}$$

and

$$\kappa_{i,2} = \kappa - \kappa_{i,1}$$

The attention allocated to a particular stock market is increasing with the asset holdings on this market since

$$\frac{\partial \kappa_{i,j}}{\partial x_{i,j}} \geq 0$$

Therefore, investors optimally decide to allocate more attention to countries where they have more asset holdings.

3 Dataset

The model presented in the previous section developed a relationship between attention allocation and asset holdings. In order to test the predictions of the model, we need to find an appropriate measure of the attention that economic agents allocate to different countries. However, this is not an easy task. First, agents process information using different media (i.e., books, magazines, newspapers, radio, TV, the internet). Second, for most of these different type of media, it is virtually impossible to track the amount of effort spent by the agent on processing information about each individual country. In this section, we show how data from internet search queries can be used to overcome these difficulties.

3.1 How to Measure Attention Allocation?

3.1.1 The World Wide Web and Internet search engines

The World Wide Web (commonly shortened to "the Web") is becoming the predominant information system. The Web is a system of interlinked hypertext documents accessible via the Internet. With a web browser, a user may access the information contained in those documents by viewing web pages and exploring their texts, images, videos, and other multimedia. However, the exponential growth in the number of documents available, one of the main reasons for its increasing popularity, also causes a problem referred to as information overload. The high rate of new documents being added to the already available large stock, combined with a significant proportion of websites with questionable reliability, make it very difficult for the user to identify what information is relevant. This is why Internet search engine tools are so popular. Search engines help minimize the time required to find reliable information and the amount of information that must be consulted.

Some numbers confirm this increasing popularity. According to the *World Development Indicators* database, the number of internet users in the US as a fraction of the total population

has increased from 43.9% in 2000 to 63.0% in 2004. And according to Fallows (2005), 84% of adult internet users have used search engines to help them find information on the Web.[1] On an average day, about 53% of internet users will go online and more than half of them will use a search engine. As the author concludes, "searching is becoming a daily habit for about a third of all internet users."

Fallows (2005) also finds that search engines are not only popular among Internet users, but also that search engine users are confident about their searching abilities, have positive experiences when using search engines and trust the search results they obtain. These conclusions were achieved after analyzing a daily tracking survey of Americans' use of the Internet. The data was collected between May 14 and June 17 of 2004, among a sample of 2,200 adults, aged 18 and older. The study's main findings show that:

- 84% of internet users have used search engines. On any given day, 56% of those online use search engines;

- 92% of those who use search engines say they are confident about their searching abilities, with over half of them, 52%, saying they're "very confident";

- 87% of searches say they have successful search experiences most of the time, including some 17% of users who say they always find the information that they are looking;

- 68% of users say that search engines are a fair and unbiased source of information; 19% say that they don't place their trust in search engines.

3.1.2 The AOL click-through series

The previous section shows that internet with the help of search engine tools is one of the most popular ways of finding information. However, we are still left with the problem of measuring attention allocated to each individual country.

Recently, AOL released the entire search/click-through record from 657,426 anonymous users collected over a three month period (01 March, 2006 - 31 May, 2006) for research purposes (Pass et al, 2007). This dataset consists of 36,389,567 lines of records, which includes

[1] Among all internet activities, only the act of sending and receiving email is more popular than using search engines.

19,442,629 user click-through records and 16,946,938 records without user click-through. Examples of the records are shown below:

142 rentdirect.com 2006-03-01 07:17:12

142 www.prescriptionfortime.com 2006-03-12 12:31:06

142 staple.com 2006-03-17 21:19:29

142 westchester.gov 2006-03-20 03:55:57 1 http://www.westchestergov.com

1337 fiserv 2006-03-24 14:05:01 1 http://www.fiserv.com

1337 fiserv 2006-03-24 14:05:01 3 http://www.fiservlendingsolutions.com

1337 fiserv 2006-03-24 14:05:01 2 http://www.fiservinsurance.com

1337 fiserv 2006-03-24 14:05:01 3 http://www.fiservlendingsolutions.com

1337 integrated real estate 2006-03-27 14:52:29 1 http://www.integratedreal.com

1337 integrated real estate 2006-03-27 14:52:29 2 http://www.irisnet.net

A click-through record corresponds to a click-through event on a result in the search result list returned by the search engine for a query. Each record contains five pieces of information (u, q, t, r, c): an anonymous user ID number (u), the query issued by the user (q), the time when the query was submitted for search (t), the rank of the clicked result on the search result page (r), and the domain portion of the clicked result URL (c). If a user clicked multiple results for a single query, it is considered as multiple click-through events and saved as multiple records.

A record without user click-through corresponds to an event that a user issued a query, but did not click on any results returned back by the search engine. The record is a triplet (u, q, t) consisting of the anonymous user ID number, the query issued by the user, and the time when the query was submitted for search.

The AOL search/click-through data help us measure the attention allocated to different countries through the Internet not only because it contains the search query, but also because it tell us whether the user clicked in any result, and what is the domain portion of the clicked result URL. While having the exact search query helps to see what information the user was looking for, many search queries are too broad or too ambiguous, yielding many results that

are irrelevant to the user. But since we know whether the user clicked through any of the results, we can identify the search results that were irrelevant. Moreover, having the web address of the results that were clicked through can help us extract additional information about the user's exact needs.

3.1.3 The Attention Allocation Variable

Using the AOL search/click-through data, we constructed the *attention* variable, which measures the number of times a user clicked through a search result from a particular country from March 1 until May 31 of 2006. This measure is constructed by taking the host name from each click-through record, and mapping it to the country where the host address has been assigned, using information from several online databases. Therefore, we are measuring the attention allocated to a country by the number of times this country provided the answer to a search query.

3.2 Other Data

We measure *holdings* using the annual survey of U.S. residents' holdings of foreign assets jointly conducted by the U.S. Treasury, the Federal Reserve Bank of New York, and the Board of Governors of the Federal Reserve System.[2] Although the asset holding values come from surveys, "*reporting to the surveys was mandatory, with penalties of noncompliance, and the data received were subjected to extensive analysis and editing*" (Ahearne, Griever and Warnock, 2004), which guarantees the accuracy of the information. The survey measured the value of U.S. holdings of foreign securities as of December 31, 2006, of approximately US$ 6 trillion, being the largest for the United Kingdom (US$ 1,076 billion), followed by Japan (US$ 596 billion) and Canada (US$ 478 billion).

The pre-determined variables are divided into three groups: the first group, $\overrightarrow{X}_1^i$, contains variables that only influence the attention allocated to each country; the second group, $\overrightarrow{X}_2^i$, includes those variables that only affect the holdings of foreign assets; and in the last group,

[2] The survey has been annual since 2003. Previous years include 1994, 1997, 1999 and 2001.

$\overrightarrow{X}_3^i$, there are variables that could potentially affect both the attention allocation and the holdings of foreign securities.

The first group includes a variable that only affect the attention allocation. This variable, *female models*, constructs a country ranking of "pop culture" interest. This ranking is based on how popular a country's female models are among the American society. More specifically, we construct this ranking using three portals about men's lifestyle: AskMen, Maxim and FHM. These portals have a target reader of 26-year-old men, with a core target of 15-34 year old, and each of them compute an annual list of the 100 most beautiful models, according to the preferences of its readers. Using the 2006 rankings, we assign $101 - \#rank^j$ points to each female model where $\#rank^j$ is her position in the ranking of magazine j (i.e. 100 points to the first place, 99 to the runner up, 98 to the third place and so on). Then, we map each female model to her country of citizenship and finally we compute the total number of points for each country based on her models' nationality. The higher a country's number of points, the more popular are its *female models* and hence the more internet searches should the country receive.

In the second group, there are the variables that only affect the holdings of foreign assets. Following Chan, Covrig and Ng (2005), we include a measure of stock market development and a measure of economic proximity. A country's stock market development is proxied by its relative size: the stock market capitalization as a percentage of GDP, *market cap*. The data is collected from the Wold Bank's *World Development Indicator* database. Since more recent data is missing for many countries, we use for each country the average stock market capitalization between 2001 and 2005. Other things equal, investors tend to prefer stock markets that have, on average, been more liquid in the past years. Economic proximity is proxied by the importance of the U.S. as a trading partner for each country i in the sample. We use the IMF's *Direction of Trade Statistics* database to calculate the size of each country i bilateral trade volume (exports and imports) with the U.S., as a percentage of country i's total trade volume in 2006. The higher the U.S. bilateral *trade importance* for country i, the more familiar we should expect them to be in terms of engaging in economic transactions with

the U.S..

The last group contains economic development and "gravity measures". Economic development captures the effect of many variables that are associated to the quality of a country's institutions. Examples of such qualities could be corruption, the confidence in the judicial system or in courts to uphold property rights, and the stability of the political system. Other things equal, we expect Americans to hold more assets in countries with more favorable investor-protection mechanisms or more stable political systems. However, economic development may also affect the level of attention allocated towards a country. More developed economies could offer websites or webservices that are either more interesting or more sophisticated, and hence attract more attention. We use the log of each country's 2006 *GDP per capita* as a measure of economic development.

Finally, "gravity measures" capture cultural differences between the U.S. and other countries in the sample. The first variable, *distance*, measures the log of the geographical distance from a country's capital city and Washington DC. The second variable, *english*, is a dummy variable that has the value of 1 for English speaking countries (which means that the country shares a common language with the U.S.) and 0 otherwise. We have no prior beliefs about how cultural differences affect the attention allocated to a country. The effect could potentially go either way. It could be positive, indicating that agents use search engines to learn about more distant cultures, or negative, suggesting that Americans are more interested in searching information about more similar cultures. With respect to asset holdings, cultural differences should affect them negatively since they might be associated to informational asymmetries, as in Portes and Rey (2006) or Ahearne, Griever and Warnock (2004). However, since the measure of attention is also related to informational asymmetries (the more attention allocated to a country, the more we learn about it, and therefore the smaller is the degree of informational asymmetry), we should not be surprised if "gravity measures" turn out to be statistically insignificant in the holdings equation.

Table 1 presents each country's descriptive statistics.

[Insert Table 1 about here]

4 Empirical Evidence

In this section, we present the empirical evidence of the two-way causal relationship between attention allocation and foreign asset holdings:

$$\begin{aligned} attention^i &= \beta_0^a + \beta_1^a holdings^i + \overrightarrow{\beta}_2^a \overrightarrow{X}_1^i + \overrightarrow{\beta}_3^a \overrightarrow{X}_3^i + \varepsilon^i \\ holdings^i &= \beta_0^h + \beta_1^h attention^i + \overrightarrow{\beta}_2^h \overrightarrow{X}_2^i + \overrightarrow{\beta}_3^h \overrightarrow{X}_3^i + \varepsilon^i \end{aligned}$$

We present the estimation output using two different type of estimation procedures, starting with equation-by-equation 2SLS and then moving to simultaneous equation 3SLS. Although the 3SLS estimator is more efficient, it requires that the system is perfectly specified. A wrong instrument in one equation would bias the coefficients of another equation in the system. In that sense, the 2SLS estimator is more robust.

4.1 Equation-by-equation 2SLS

Table 2 presents the estimation output obtained through individual 2SLS regressions.[3] The first column of results refers to the *attention* equation. In this equation, we used *market cap* and *trade importance* as the instrumental variables for *holdings*. We can see that the overidenfication test does not reject the validity of the chosen instruments, the Hansen-Sargan statistic is 0.372, with a p-value of 54.2%. The model also has a good in-sample fit, the R^2 is 77.1%, and all estimated coefficients are significant at the 1% significance level. First, agents do allocate more attention to countries where they hold more assets. A 1% increase in the *holdings* of a particular country's asset is also associated with a 0.25% increase in the number of that country's click-throughs. Second, *attention* is affected by a country's popularity among internet users: an increase of 100 points in the *female model* popularity rank increases the number of click-throughs received by a country by 0.40%. Third, the coefficients associated to the "gravitational" variables show that AOL users tend to search more information about

[3] Table 4 presents the estimation results of the first stage regressions. We can see that the pre-determined variables can explain 67.0% and 63.1% of the cross-country variations in *attention* allocation and international *holdings*.

more distant cultures: a 1% increase in the geographical distance between a country and the U.S. increases the number of click-throughs by 0.76%; and English speaking countries, other things equal, receive 0.82% less click-throughs compared to non-English speaking countries. Finally, the higher is the level of development of a country, the more attention receives: a 1% increase in *GDP per capita* increases the number of click-throughs by 0.59%.

[Insert Table 2 about here]

The second column of Table 2 presents the estimated coefficients for the *holdings* equation. None of the two "gravitational" variables (*distance* and *english*) are included as regressors since they were not statistically significant.[4] Since we have seen that both gravitational variables were significant in the *attention* equation, both of them will join the *female models* ranking as instruments for *attention* allocation in the *holdings* equation. The overidenfication test does not reject the validity of these three variables as instruments: the Hansen-Sargan statistic is 3.208, with a p-value of 20.1%. The in-sample fit is also good, with an R^2 of 73.6%. The coefficient associated to *attention* is significant at the 1% significance level: countries that receive 1% more click-throughs also receive 0.98% more asset holdings. Second, investors tend to invest more in more developed stock markets: a 1% increase in the *market cap* as a percentage of GDP increases the asset holdings by 0.53%, with the coefficient being significant at the 5% level. Third, the coefficient that captures economic familiarity is also significant at the 1% level: a 1% increase in the bilateral *trade importance* with the U.S. increases asset holdings by 0.75%. Finally, economic development also has a positive and significant effect (at the 10% level) on international investment: a 1% increase in *GDP per capita* increases asset holdings by 0.47%.

[4] Table 5 presents the estimation output of the *holdings* equation when both *english* and *distance* are included as regressors. We can see from column (1) that while *distance* is significant, *english* is not. Thefore, we reestimate the holdings equation excluding *english* as one of the regressors. We can see from the results of the reestimated system presented in column (2) that *distance* is not statistically significant anymore. Therefore, we excluded both variables from the list of regressors in the *holdings* equation. This means that *attention* is probably already capturing the effect of informational asymmetries on the foreign asset holdings.

4.2 Multiple-equation 3SLS

The results of the 3SLS estimation are reported on Table 3. Once again, the gravitational variables were not included in the *holdings* equation since they were not significant.[5] All coefficients have the same sign and very similar magnitudes relative to the ones estimated by 2SLS. An important difference is the change in the significance level of *market cap* and *GDP per capita*. The coefficient associated to *market cap* becomes significant at the 1% level (it was significant at the 5% significance level in the 2SLS estimation). *GDP per capita* has its significance level changed from 1% to 5% in the *attention* equation and becomes not significant in the *holdings* equation.

[Insert Table 3 about here]

Since the system is over-identified, we can test the null hypothesis that all the instruments are valid. The Hansen-Sargan statistic obtained was 4.004, with a p-value of 0.261. This implies that we cannot reject the orthogonality conditions assumed when estimating the system by 3SLS.

5 Conclusion

This paper finds empirical evidence on the interaction between international asset holdings and attention allocation. We analyze a new a database that includes over 21 million web queries from 657,426 America Online users. The data includes all searches and results from those users for a three month period as well as whether they clicked or did not click on a search result and where it appeared on the result page. We explore the AOL dataset to help us measure the attention allocated to different countries. We believe that a measure based on the internet search queries can proxy attention allocation because: first, the World Wide Web is becoming the predominant information media; second, search engines are the most popular tool to help users find reliable information on the "Web" since it minimizes the time required

[5] Table 6 presents the 3SLS estimation output of the system with the "gravitational variables". If both *english* and *distance* are included, english is not significant. When we reestimate the system without *english* as one of the regressors for *holdings*, *distance* is no longer significant.

and the amount of information which must be consulted; third, by having the exact search query we know what was the topic of interest of the user.

Using the AOL dataset, we constructed an attention allocation variable, which measures the number of times a user clicked through a search result from a particular country. We combine the click-through series with data from the latest survey of U.S. portfolio holdings of foreign securities. We find empirical evidence using both 2SLS and 3SLS that the strong relationship between asset holdings and attention allocation presented by Figure 1 is actually a result of three factors. First, correlation caused by the level of development, which is captured by the positive effect of GDP per capita in both equations. Investors hold more assets in countries with more favorable investor-protection mechanisms and more stable political systems, and they also search more information from developed economies due to their more sophisticated websites or webservices. Second, causality runs from asset holdings to attention allocation. Using instrumental variables that capture the level of development in financial markets, the economic proximity and the importance of the U.S. as a trade partner for that country, we find that agents allocate more attention to countries where they hold more assets. Third, causality also runs from attention allocation to asset holdings. Using instrumental variables that are related to a country's popularity among internet users, we find that international investors tend to hold more assets in countries where they process more information.

References

[1] Ahearne, A., W. Griever, F. Warnock (2004). Information costs and home bias: An analysis of U.S. holdings of foreign equities. Journal of International Economics 62, pp. 313-336.

[2] Bacchetta, P., E. Van Wincoop (2007). Incomplete Information Processing: A Solution to the Forward Discount Puzzle. Working paper.

[3] Chan, K., V. Covrig, L. Ng (2005). What determines the domestic bias and foreign bias?. Journal of Finance 60, pp. 1495-1534.

[4] Falkinger, J. (2007). Attention Economies. Journal of Economic Theory 133, pp. 266-294.

[5] Fallows, D. (2005). Search Engine Users. Washington, DC: Pew Internet & American Life Project.

[6] Gabaix, X. and D. Laibson (2002). The 6D Bias and the Equity Premium Puzzle. In B. Bernanke, K. Rogoff (Eds.), NBER Macroeconomic Annual, pp. 257-312.

[7] Gabaix, X., D. Laibson and G. Moloche (2003). The Allocation of Attention: Theory and Evidence. Working Paper.

[8] Joachims, T. (2002). Optimizing Search Engines Using Clickthrough Data, Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), ACM.

[9] Luo, Y. (2007). Consumption Dynamics under Information Processing Constraints. Working paper.

[10] Mackoviak, B., M. Wiederholt (2007). Optimal Sticky Prices under Rational Inattention. Working paper.

[11] Mankiw, G., R. Reis (2002). Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. Quarterly Journal of Economics 117, pp. 1295-1328.

[12] Mondria, J. (2007). Financial Contagion and Attention Allocation. Working paper.

[13] Mondria, J., C. Quintana-Domeque (2006). Financial Contagion through Attention Reallocation: An Empirical Analysis. Working paper.

[14] Pass, G., A. Chowdhury, C. Torgeson, A Picture of Search. The First International Conference on Scalable Information Systems, Hong Kong, June.

[15] Portes, R. and H. Rey (2005). The Determinants of Cross Border Equity Transacton Flows. Journal of International Economics 65, pp. 269-296.

[16] Sims, C. (2006). Rational Inattention: A Research Agenda. The American Economic Review, 96, pp. 158-163.

[17] Van Nieuwerburgh, S., L. Veldkamp (2008a). Information Acquisition and Portfolio Under-Diversification. Working paper.

[18] Van Nieuwerburgh, S., L. Veldkamp (2008b). Information Immobility and the Home Bias Puzzle. Forthcoming in Journal of Finance.

Figure 1: AOL Attention Allocation and International Investment

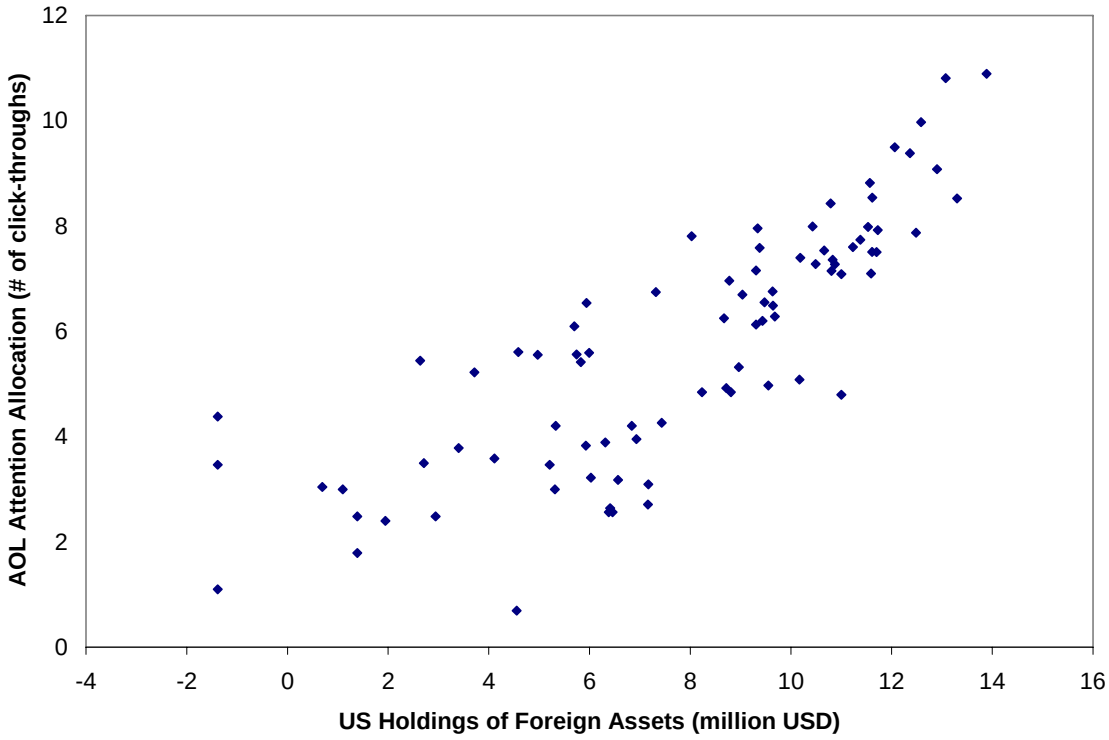

Table 1: Country Statistics

Country Name	AOL Attention Allocation Mar-May 2006	US Holdings of Foreign Assets Dec 2006	Female Model Popularity Rank 2006	Market Capitalization (% of GDP) 2001-2005 average
Argentina	1,281	11,019	0	53.39
Australia	13,331	173,153	228	104.70
Austria	1,632	26,568	0	23.95
Bangladesh	12	4	0	3.80
Belgium	1,447	35,968	0	96.25
Bolivia	32	0	0	19.90
Botswana	6	4	0	25.01
Brazil	1,824	110,294	412	40.84
Bulgaria	260	312	0	9.46
Canada	49,480	477,888	1,197	107.39
Chile	492	12,539	0	101.28
China	1,998	75,314	37	36.17
Colombia	517	5,827	190	21.82
Costa Rica	225	340	0	12.31
Cote d'Ivoire	2	95	0	12.51
Croatia	268	400	0	23.72
Czech Republic	2,462	3,070	232	23.00
Denmark	2,957	33,978	0	57.34
Ecuador	49	552	0	7.59
Egypt, Arab Rep.	127	6,685	0	44.99
El Salvador	67	935	0	14.46
Estonia	272	98	0	35.07
Fiji	11	7	0	18.09
Finland	1,196	59,934	0	114.58
France	8,765	401,388	81	79.98
Georgia	44	30	0	4.10
Germany	21,432	292,103	197	44.44
Ghana	20	3	0	16.65
Greece	535	16,012	0	61.74
Guatemala	67	206	0	1.11
Hong Kong, China	2,291	87,518	0	424.63
Hungary	810	8,409	0	23.31
Iceland	204	7,829	0	104.47
India	1,271	49,231	54	43.93
Indonesia	144	14,072	0	21.97
Ireland	1,816	120,513	0	58.72
Italy	6,759	105,893	95	43.70
Jamaica	13	591	0	104.72
Japan	5,031	596,239	26	73.22
Jordan	36	61	0	141.85
Kazakhstan	15	1,288	0	9.26
Kenya	33	15	0	21.00
Korea, Rep.	2,762	123,876	22	59.85
Latvia	231	14	0	10.81
Lebanon	46	375	52	11.20

(*continued*)

Table 1: Country Statistics (continued)

Country Name	AOL Attention Allocation Mar-May 2006	US Holdings of Foreign Assets Dec 2006	Female Model Popularity Rank 2006	Market Capitalization (% of GDP) 2001-2005 average
Lithuania	185	41	0	19.95
Luxembourg	121	60,101	0	129.14
Malaysia	658	15,404	0	145.49
Mauritania	3	0	0	97.19
Mauritius	13	639	0	34.14
Mexico	1,208	108,450	208	22.33
Moldova	80	0	0	23.73
Morocco	25	415	0	36.69
Namibia	21	2	0	6.30
Netherlands	11,868	234,066	0	103.21
New Zealand	2,853	11,372	71	38.45
Norway	1,571	50,573	31	47.74
Pakistan	52	1,026	0	22.44
Panama	161	26,069	0	25.32
Papua New Guinea	22	1,298	0	67.34
Peru	127	3,763	0	28.88
Philippines	458	10,989	0	31.55
Poland	1,966	11,816	40	20.89
Portugal	1,054	6,485	0	37.82
Romania	691	380	0	12.21
Russian Federation	4,571	48,441	194	46.27
Singapore	1,443	52,731	0	149.84
Slovak Republic	443	299	0	8.67
Slovenia	258	144	33	22.62
South Africa	1,870	42,686	298	177.83
Spain	5,107	110,957	178	80.60
Sri Lanka	32	183	0	15.20
Sweden	2,934	102,066	83	99.25
Switzerland	2,625	264,243	0	232.26
Thailand	700	13,047	0	58.59
Trinidad and Tobago	24	711	0	94.57
Tunisia	14	607	0	10.18
Turkey	859	15,284	0	31.26
Ukraine	851	1,506	54	13.34
United Kingdom	53,856	1,075,579	1,195	135.35
Uruguay	71	1,694	0	1.72
Venezuela, RB	137	6,097	0	4.56
Zambia	12	19	0	8.29
Zimbabwe	20	203	0	65.57
Mean	2,753	61,191	85	54.46
Median	358	6,291	0	35
Standard Deviation	8,342	153,410	286	62
Minimum	2	0	0	1
Maximum	53,856	1,075,579	2,006	425

(*continued*)

Table 1: Country Statistics (continued)

Country Name	Bilateral Trade with U.S. (% of partner's GDP) 2006	Distance to U.S. (miles) -	English Speaking Dummy -	Per capita GDP (current USD) 2006
Argentina	11.19	5,187	0	5,472
Australia	10.22	9,929	1	37,434
Austria	4.40	4,446	0	39,131
Bangladesh	13.90	8,106	0	429
Belgium	5.48	3,894	0	37,384
Bolivia	10.23	3,847	0	1,195
Botswana	7.58	7,937	1	5,875
Brazil	19.11	4,192	0	5,660
Bulgaria	2.19	4,958	0	4,089
Canada	74.00	456	1	38,440
Chile	18.17	4,997	0	8,865
China	18.80	7,006	0	2,034
Colombia	32.27	2,374	0	2,982
Costa Rica	33.24	2,047	0	5,047
Cote d'Ivoire	7.46	4,984	0	947
Croatia	1.77	4,536	0	9,612
Czech Republic	2.04	4,305	0	13,877
Denmark	4.76	4,075	0	50,702
Ecuador	41.25	2,690	0	3,042
Egypt, Arab Rep.	12.37	5,833	0	1,426
El Salvador	38.61	1,886	0	2,618
Estonia	3.12	4,353	0	12,237
Fiji	9.18	7,781	1	3,306
Finland	5.02	4,324	0	39,856
France	5.95	3,862	0	36,547
Georgia	6.71	5,882	0	1,702
Germany	6.73	4,011	0	35,270
Ghana	5.64	5,325	1	573
Greece	3.17	5,141	0	22,042
Guatemala	39.87	1,866	0	2,735
Hong Kong, China	5.92	8,144	1	27,072
Hungary	2.68	4,594	0	11,227
Iceland	7.07	2,822	0	53,029
India	12.87	7,548	1	817
Indonesia	8.37	10,187	0	1,634
Ireland	17.46	3,395	1	52,893
Italy	5.58	4,535	0	31,496
Jamaica	34.48	1,444	1	3,954
Japan	17.35	6,834	0	34,023
Jordan	12.14	5,985	0	2,538
Kazakhstan	2.99	6,572	0	5,045
Kenya	7.68	7,554	1	603
Korea, Rep.	12.57	6,994	0	18,341
Latvia	2.79	4,444	0	8,797
Lebanon	8.19	5,868	0	5,603

(*continued*)

Table 1: Country Statistics (continued)

Country Name	Bilateral Trade with U.S. (% of partner's GDP) 2006	Distance to U.S. (miles) -	English Speaking Dummy -	Per capita GDP (current USD) 2006
Lithuania	4.02	4,564	0	8,770
Luxembourg	2.49	3,979	0	89,564
Malaysia	15.14	9,571	0	5,780
Mauritania	5.31	3,909	0	844
Mauritius	4.86	9,481	1	5,145
Mexico	72.23	1,885	0	8,052
Moldova	1.43	4,969	0	850
Morocco	3.89	3,858	0	1,879
Namibia	8.86	7,394	1	3,107
Netherlands	5.70	3,865	0	40,167
New Zealand	13.08	8,765	1	25,179
Norway	5.29	3,903	0	66,964
Pakistan	12.45	7,128	1	810
Panama	10.03	2,073	0	5,205
Papua New Guinea	2.09	9,039	1	943
Peru	23.98	3,523	0	3,288
Philippines	13.93	8,626	1	1,382
Poland	1.89	4,488	0	8,888
Portugal	4.08	3,591	0	18,185
Romania	2.00	4,992	0	5,645
Russian Federation	5.27	4,909	0	6,932
Singapore	10.52	9,709	1	30,082
Slovak Republic	2.37	4,479	0	10,223
Slovenia	1.86	4,441	0	18,674
South Africa	9.27	8,099	1	5,381
Spain	3.33	3,806	0	28,108
Sri Lanka	13.79	9,022	1	1,364
Sweden	6.72	4,145	0	42,553
Switzerland	9.08	4,134	0	51,033
Thailand	12.11	8,871	0	3,187
Trinidad and Tobago	52.41	2,179	1	15,214
Tunisia	3.26	4,600	0	2,990
Turkey	5.37	5,476	0	5,521
Ukraine	2.63	4,900	0	2,278
United Kingdom	10.59	3,663	1	38,850
Uruguay	10.29	5,226	0	5,828
Venezuela, RB	45.65	2,088	0	6,730
Zambia	1.41	7,716	1	919
Zimbabwe	3.90	7,932	1	383
Mean	35.78	5,264	0.27	14,900
Median	7.58	4,597	0	5,660
Standard Deviation	216.72	2,266	0.45	18,244
Minimum	1.41	456	0	383
Maximum	2006.00	10,187	1	89,564

Table 2: Two Stages Least Squares Estimation Output – Second Stage

		(1)	(2)
Estimation method:		2SLS	2SLS
Dependent variable:		*attention*	*holdings*
Regressors:	*attention*	-	0.981***
		-	(0.169)
	holdings	0.254***	-
		(0.080)	-
	female models	0.004***	-
		(0.001)	-
	market cap	-	0.527**
		-	(0.223)
	trade importance	-	0.753***
		-	(0.227)
	distance	0.764***	-
		(0.247)	-
	english	-0.822***	-
		(0.318)	-
	GDP per capita	0.585***	0.466*
		(0.166)	(0.262)
Overidentification test			
Hansen-Sargan statistic		0.372	3.208
		(0.542)	(0.201)
Number of observations		84	84
R^2		77.1%	73.6%

Note: Each equation was estimated individually by two stages least squares. White's robust standard errors are given in parenthesis under the coefficients and and p-values are given in parenthesis under the Hansen-Sargan statistic. The symbols ***, **, and * denote that the individual coefficient is significant at the 1%, 5%, and 10% significance level respectively.
level respectively.

Table 3: Three Stages Least Squares Estimation Output

Estimation method:		**3SLS**	
Dependent variable:		*attention*	*holdings*
Regressors:	*attention*	-	0.977***
		-	(0.229)
	holdings	0.286***	-
		(0.092)	-
	female models	0.004***	-
		(0.001)	-
	market cap	-	0.552***
		-	(0.216)
	trade importance	-	0.737***
		-	(0.221)
	distance	0.864***	-
		(0.283)	-
	english	-0.897***	-
		(0.295)	-
	GDP per capita	0.545**	0.460
		(0.178)	(0.289)
Overidentification test			
Hansen-Sargan statistic		4.004	
		(0.261)	
Number of observations		84	
R^2		77.3%	73.6%

Note: Both equations were estimated simultaneously by three stages least squares Standard errors are given in parenthesis under the coefficients, and the p-value is given in parenthesis under the Hansen-Sargan statistic. The symbols ***, **, and * denote that the individual coefficient is significant at the 1%, 5%, and 10% significance level respectively.

Table 4: Two Stages Least Squares Estimation Output – Fist Stage

		(1)	(2)
Estimation method:		OLS	OLS
Dependent variable:		*attention*	*holdings*
Regressors:	*female models*	0.005***	0.005***
		(0.001)	(0.001)
	market cap	0.262	0.766**
		(0.169)	(0.352)
	trade importance	0.271*	1.254***
		(0.152)	(0.302)
	distance	1.193***	2.074***
		(0.322)	(0.588)
	english	-1.277***	-1.709**
		(0.409)	(0.750)
	GDP per capita	0.922***	1.497***
		(0.153)	(0.312)
Number of observations		84	86
R^2		67.0%	63.1%

Note: Each equation was estimated individually by ordinary least squares. White's robust standard errors are given in parenthesis under the coefficients. The symbols ***, **, and * denote that the individual coefficient is significant at the 1%, 5%, and 10% significance level respectively.

Table 5: Two Stages Least Squares Estimation Output – Second Stage

		(1)	(2)
Estimation method:		2SLS	2SLS
Dependent variable:		*holdings*	*holdings*
Regressors:	*attention*	0.878***	0.944***
		(0.112)	(0.150)
	holdings	-	-
		-	-
	female models	-	-
		-	-
	market cap	0.548**	0.452*
		(0.272)	(0.235)
	trade importance	1.034***	0.937***
		(0.266)	(0.273)
	distance	0.926**	0.752
		(0.431)	(0.518)
	english	-0.643	-
		(0.490)	-
	GDP per capita	0.641**	0.616**
		(0.276)	(0.294)
Overidentification test			
Hansen-Sargan statistic		-	1.592
		-	(0.207)
Number of observations		84	84
R^2		74.7%	74.3%

Note: Each equation was estimated individually by two stages least squares. White's robust standard errors are given in parenthesis under the coefficients and and p-values are given in parenthesis under the Hansen-Sargan statistic. The symbols ***, **, and * denote that the individual coefficient is significant at the 1%, 5%, and 10% significance level respectively.
level respectively.

Table 6: Three Stages Least Squares Estimation Output

		(1)		(2)	
Estimation method:		**3SLS**		**3SLS**	
Dependent variable:		*attention*	*holdings*	*attention*	*holdings*
Regressors:	*attention*	-	0.874***	-	0.940***
		-	(0.232)	-	(0.227)
	holdings	0.254***	-	0.272***	-
		(0.094)	-	(0.092)	-
	female models	0.004***	-	0.004***	-
		(0.001)	-	(0.001)	-
	market cap	-	0.590**	-	0.499**
		-	(0.231)	-	(0.217)
	trade importance	-	1.003***	-	0.899***
		-	(0.256)	-	(0.243)
	distance	0.764***	0.887*	0.756***	0.702
		(0.293)	(0.487)	(0.293)	(0.467)
	english	-0.822***	-0.653	-0.924***	-
		(0.308)	(0.524)	(0.297)	-
	GDP per capita	0.585***	0.625**	0.553***	0.597**
		(0.180)	(0.298)	(0.178)	(0.299)
Overidentification test					
Hansen-Sargan statistic		0.328		1.866	
		(0.567)		(0.393)	
Number of observations		84		84	
R^2		77.1%	74.7%	77.2%	74.3%

Note: Each system of two equations were estimated simultaneously by three stages least squares. Standard errors are given in parenthesis under the coefficients, and the p-value is given in parenthesis under the Hansen-Sargan statistic. The symbols ***, **, and * denote that the individual coefficient is significant at the 1%, 5%, and 10% significance level respectively.