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

Jordi Mondria[†]
University of Toronto

Thomas Wu[‡]
UC Santa Cruz

Yi Zhang[§]
UC Santa Cruz

April 2010

Abstract

This paper explores the joint determination of home bias and attention allocation. We overcome the typical challenge associated with evaluating attention allocation theories by using a new internet search query dataset to measure how much information investors decide to process. Employing an instrumental variables approach, we find empirical evidence of a two-way causality between home bias and attention. Our estimates suggest that if all countries were to receive the same level of attention as the U.S., then the average home bias by U.S. investors would fall from 85.2% to 57.3%.

Keywords: Home Bias, Asymmetric Information, Attention Allocation , Internet Search Query.

JEL Codes: F30, D82, G11.

1 Introduction

The existence of a domestic bias in international equity holdings was first documented by French and Poterba (1991) and Tesar and Werner (1995). One popular justification focuses on the role of informational asymmetries between domestic and foreign investors. Although over the past decade innovations in telecommunication and information technologies have greatly increased investors' access to information around the globe, significant levels of home bias still remain. To explain this persistent bias, information-based models commonly have to assume implausibly large exogenous information asymmetries. Recent work by Van Nieuwerburgh and Veldkamp (2009), however, generated these large information asymmetries by endogeneizing the investors' information structure. In their model, the allocation of investors' limited attention leads to information heterogeneity endogenously and thereby a sizeable bias in portfolios. Our paper provides empirical support for attention-based explanations of the home equity bias.

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.[1] Although attention allocation models are very appealing from a theoretical perspective, empirical evaluation remains a challenge. The main difficulty relies on measuring the pieces of information each agent processes. Our paper overcomes such a challenge by using a new dataset from America Online (henceforth AOL) that includes over 21 million web searches by 657,426 customers. We then combine data on U.S. holdings of foreign securities to analyze the joint determination of home bias and attention allocation by U.S. investors.

Our contributions are three-fold. First, we derive a simple model based on Van Nieuwer-

---

[1]In macroeconomics, inattentiveness has been used by Mackoviak and Wiederholt (2009a) and Mankiw and Reis (2002) to explain sticky prices and by Luo (2008), Mackoviak and Wiederholt (2009b) and Tutino (2009) to explain business cycle dynamics. 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), price comovement in Mondria (2010), portfolio under-diversification in Van Nieuwerburgh and Veldkamp (2010), home equity bias in Van Nieuwerburgh and Veldkamp (2009) and the forward discount bias in Bacchetta and Van Wincoop (2009).

burgh and Veldkamp (2009) to study the channels through which attention allocation affects portfolio choice and vice versa. Second, we provide a new measure of the attention allocated across countries by U.S. investors. Measuring attention allocation is a complicated task. Agents process information using different media (i.e., books, magazines, newspapers, radio, TV, the Internet) and, for most of these media, it is virtually impossible to track the amount of effort spent processing information about different topics. On August 3, 2006, AOL released a dataset that includes all web search queries from its 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. A web search query is the exact phrase that a user types into the search engine to satisfy his or her information needs. 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 internet search queries can proxy attention allocation because: (i) the World Wide Web is becoming the predominant information medium; (ii) search engines are the most popular tool for finding reliable information on the "Web" since they minimize the time required and the amount of information which must be consulted; (iii) by having user's exact search query, we know the topic that he or she was interested in. Using this dataset, we construct a click-through series that counts the number of times a user clicked on 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.

Our third contribution is to study the determinants of attention allocation and home equity bias across countries by U.S. investors. We combine the click-through series with home bias data constructed from the 2006 survey of U.S. portfolio holdings of foreign securities. We recognize that both home bias and attention allocation are endogenous and, using instrumental variables as our estimation methodology, we find empirical evidence of joint causality. First, causality runs from asset holdings to attention allocation. Using instrumental variables that capture implicit financial costs, we find that agents do indeed allocate more attention to countries whose assets make up a greater share of their portfolios. Second, causality also runs from attention allocation to asset holdings. Using instrumental variables that are related

2

to a country's popularity among internet users, we find that international investors favor assets from more familiar countries. Taken together, our results show that the unconditional correlation between measures of information and equity holdings is not just due to agents being more interested in countries whose assets they already hold. Instead, agents endogenously increase their holdings of a particular country's assets in response to an exogenous increase in the information they have about that country. Our results also suggest that if all countries were to receive the same level of attention as the U.S., then the average home bias by U.S. investors would fall from 85.2% to 57.3%.

Many studies have emphasized the importance of asymmetric information in determining home bias.[2] Ahearne, Griever and Warnock (2004), for example, provide evidence that information regarding accounting standards and practices is an important determinant of the distribution of home bias across countries. Dahlquist, Pinkowitz, Stulz and Williamson (2003) discuss the importance of information about cross-country differences in corporate governance while Kraay, Loayza, Serven and Ventura (2005) suggest that information regarding sovereign risks and political events helps explain differences in country portfolios. Bekaert (1995) presents evidence that information about macroeconomic conditions, market structure and institutions are important drivers of equity flows into emerging markets and Portes and Rey (2005) and Chan, Covrig and Ng (2005) find that the geography of information is relevant for explaining cross-border equity flows and mutual fund equity holdings respectively. Moreover, Coval and Moskowitz (2001) and Malloy (2005) show that fund managers are able to predict abnormal returns in local investments. Because these results are stronger for investments in small firms and firms located in remote areas, the authors argue that local investments allow fund managers to assess and survey the firm's operations and have regular meetings with CEOs. This evidence indicates that information about firm operations and close contact with CEOs generates local bias at home.[3]

---

[2]For theoretical work on the home equity bias puzzle and asymmetric information, see Gehrig (1993), Brennan and Cao (1997), Zhou (1998), Barron and Ni (2008), Hatchondo (2008) and Van Nieuwerburgh and Veldkamp (2009).

[3]However, the evidence provided by Kang and Stulz (1997) and Dahlquist and Robertsson (2001), who show that foreign investors hold stock of firms that are large and have low leverage, suggests that the effect of this information on investments across the border might be rather small.

The rest of the paper is organized as follows. Section 2 develops the model. Section 3 describes the new measure of attention allocation. Section 4 presents the dataset. Section 5 describes the empirical methodology. Section 6 shows the empirical results. Section 7 concludes.

2　Model

The objective of the model is to understand how attention allocation affects asset holdings and how asset holdings affect attention allocation.

2.1　Model Description

We formulate a partial equilibrium model based on Van Nieuwerburgh and Veldkamp (2009, 2010) with a continuum of investors and two countries.[4] 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 and, based on this allocation, obtain private information about each market, $\tilde{Y}_i = (\tilde{y}_{i,1}, \tilde{y}_{i,2})'$. Third, each investor chooses optimal asset holdings $X_i = (x_{i,1}, x_{i,2})'$ given $\tilde{Y}_i$. Fourth, trading ceases and investors consume the payoffs realized by their portfolios.

Investors have an absolute risk tolerance parameter ρ and 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

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

---

[4]For a full general equilibrium model, where attention allocation and asset holdings are jointly determined, see Van Nieuwerburgh and Veldkamp (2009) and Mondria and Wu (2010).

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, $\tilde{R} = (\tilde{r}_1, \tilde{r}_2)'$ with $\tilde{r}_j \sim N\left(\bar{r}_j, \sigma_{r,j}^2\right)$ for $j = 1, 2$ is the vector of risky asset payoffs, $\tilde{Y}_i = (\tilde{y}_{i,1}, \tilde{y}_{i,2})'$ is agent i's 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 $\tilde{R}$. Agent 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 for processing information about the state of the economy[5]

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

where $\kappa_{i,j}$ is the attention allocated by agent i to stock market j. In other words, the precision of a private signal is higher if more attention is allocated to that particular signal. However, investors are constrained in their capacity to process information about asset payoffs:

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

The information constraint imposes a limit on the reduction of agent i's uncertainty about the future payoff of risky asset j so the information processing resources have to be optimally divided between the two risky assets. After deciding how much information to process about each stock market, investors use Bayesian updating to incorporate the private signal, $\tilde{Y}_i$, into their beliefs about asset payoffs and, based on these posterior beliefs, they then choose their optimal asset holdings.

---

[5] Peng and Xiong (2006) assume an increasing returns to scale technology to process information, which implies $\sigma_{ij}^2 = \frac{\sigma_{rj}^2}{(e^{\kappa_{ij}} - 1)}$. However, as in Van Nieuwerburgh and Veldkamp (2010), we assume a linear technology that implies constant returns to scale.

2.2 Effect of Attention Allocation on Asset Holdings

Suppose each agent chooses risky asset holdings taking the attention allocation as given. After observing a private signal about each risky asset, investors derive posterior beliefs about asset payoffs in order to determine their optimal 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

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

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

2.3 Effect of Asset Holdings on Attention Allocation

Suppose that investors take as given their asset holdings when choosing where to allocate information resources. The objective function is then

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

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 increases with the asset holdings from that market since

$$\frac{\partial \kappa_{i,j}}{\partial x_{i,j}} \geq 0 \tag{2}$$

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

3 Measuring Attention Allocation

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 allocated to different countries by economic agents. However, this is not an easy task. First, agents process information using various media (i.e., books, magazines, newspapers, radio, TV, the Internet). Second, for most of these media, it is virtually impossible to track the amount of effort each agent spends on processing information about individual countries. In this section, we show how data from Internet search queries help us overcome these difficulties.

3.1 The World Wide Web and Internet search engines

The World Wide Web (commonly shortened to "the Web") is becoming a predominant information system. The Web is a system of interlinked hypertext documents accessible via the

Internet. With a web browser, a user may access information contained in these 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 creates a problem referred to as information overload. The high rate of new documents being added to the already available and large stock, combined with a significant proportion of websites with questionable reliability, make it very difficult for the user to identify relevant information. For this reason, Internet search engine tools are very popular. Search engines help minimize the time required to locate reliable information and the amount of information that is consulted.

Some numbers confirm this increasing popularity. According to the *World Development Indicators* database, the number of Internet users in the U.S. as a fraction of the total population increased from 43.9% in 2000 to 63.0% in 2004. Moreover, according to Fallows (2005), 84% of adult Internet users have used search engines to help them find information on the Web.[6] On an average day, about 53% of internet users 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."

In addition to the popularity of search engines, Fallows (2005) also finds 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 reached after analyzing a daily tracking survey of the Internet use of 2,200 American adults, aged 18 and older, over the period May 14 - June 17, 2004.

3.2   The AOL click-through series

The previous section established that the 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 the attention allocated to each individual country.

Recently, AOL released the search/click-through histories of 657,426 anonymous users over

---

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

a three month period (01 March, 2006 - 31 May, 2006) for research purposes (Pass et al, 2006). This dataset consists of 36,389,567 records, including 19,442,629 user click-through records and 16,946,938 records without user click-through information. 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 is registered when a user clicks on a result from the list returned by the search engine for his or her query. Each record contains five pieces of information (u, q, t, r, c): an anonymous user ID (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's URL (c). If a user clicks several results for a single query, each click is recorded separately.

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

The AOL search/click-through data helps us measure the attention allocated to different countries through the Internet not only because it contains the search query, but also because it tells us whether the user clicked on any of the returned results and, if so, the domain of the clicked URL. While having the exact search query helps us see the information the user

was after, many search queries tend to be short and ambiguous so not all search results are relevant to the user. However, by knowing whether the user clicked through any of the results, we can identify the relevant ones. Moreover, by having the web address of the results that were clicked through, we can extract additional information about the user's exact needs.

4   Data

In this section, we present the variables used in our empirical analysis. We construct a quantitative measure of attention allocation from the AOL click-through series. We also describe how we combine data on U.S. holdings of foreign equities with data on market capitalization to calculate home bias. Finally, we describe additional variables that we use as instruments and controls. All data refers to the year 2006, unless otherwise noted.

4.1   Attention Allocation

Using the AOL search/click-through data, we construct an *attention* variable, which measures the number of times a user clicked through a search result from a particular country between March 1 and May 31 of 2006. In other words, we are measuring the attention allocated to a country by the number of times that country provided the answer to a search query. 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. This procedure first involves finding the IP address associated with the domain portion of the website name (i.e, DNS lookup). Since IP addresses are allocated by regional organizations, it is then relatively simple to find the country associated with the IP address. Table 1 lists the *attention* level each country in our sample receives from AOL users. We can see that the Top 5 countries are United Kingdom, Canada, Germany, Netherlands and Australia.

[**Insert Table 1 about here**]

Different types of investors possess different information sets, as each requires distinct information to evaluate financial assets. We believe the related existing literature has cap-

10

tured information gathering by fund managers. Institutional investors are mostly influenced by information regarding fundamentals and institutions in a broad sense. Our web search query data represents the information processing of individual investors and the cross-country networks associated with migration and cultural ties. The attention allocated by individual investors across countries does not only affect their investment strategies, but also institutional investors' strategies. As argued by Chan, Covrig and Ng (2005), fund managers' investment strategies are directly influenced by the information and the preferences of their clients, who primarily are individual investors.

4.2 Home Bias

We combine market capitalization data from the World Bank's *World Development Indicators* database with cross-border portfolio holdings data from the U.S. Treasury *International Capital System* to calculate a measure of U.S. investors' *home bias* towards countries in our sample. Following Ahearne et al (2004) methodology, we calculate each country's share in the U.S. equity portfolio, and then we normalize the value by the country's share in the world portfolio. The resulting variable measures the degree in which the amount of U.S. equity holdings in each country deviates from those predicted by ICAPM. We can see in Table 1 that the five countries towards which U.S. investors have the least home bias are Ireland, Mexico, Switzerland, Finland, and Netherlands.

4.3 Instrumental Variables

We will use three instruments for *attention* allocation. Our first instrument is an indicator of *cultural sites* within a country. We use the number of World Heritage *cultural sites* from the *UNESCO/World Heritage Centre list*. Our second instrument is a *female models ranking* among adult males in the U.S.. We construct this ranking using three portals regarding men's lifestyle: AskMen, Maxim and FHM. These portals have a target audience of 15-34 years old men. 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. Finally, we compute the total number of points for each country based on the 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. Table 1 also shows that the five countries with the highest number of points awarded to its *female models* are Canada, United Kingdom, Brazil, South Africa, and Czech Republic. Our third instrument for attention allocation is the number of *Internet users* (per 100) obtained from the Telecommunication Union *World Telecommunication Development Report and Database.* We will discuss the validity of these instruments in the following section.

We will use two instruments for *home bias*. Our first instrument is a qualitative measure of each country's financial *market sophistication* level. The data is collected from the World Economic Forum's annual *Global Competitiveness Report* and ranges from 1 (poor by international standards) to 7 (excellent by international standards). Our second instrument is the number of procedures required to *enforce a contract* in each country, collected from the World Bank's *World Development Indicators* database. The higher the number of procedures in a given country, the lower the quality of the regulatory framework. We will discuss the validity of the above instruments in the following section.

4.4   Additional Controls

We also include additional control variables that may potentially affect *attention* or *home bias*. First we consider two measures of country size: total *population* and total *land area* in square kilometers. It is natural to expect a larger country to receive more *attention*; however, it is unclear whether U.S. investors have a smaller *home bias* towards larger countries. Since the *home bias* variable measures the deviation of actual U.S. holdings relative to those predicted by ICAPM, it already incorporates the country size.

Following, we look at measures of economic development. Since more developed economies tend to provide better services, they may receive more *attention* and less *home bias* from U.S.

investors. Our first proxy for development level is *GDP per capita*. Our second proxy is a dummy variable that equals 1 for all countries that were included in the International Monetary Fund's list of *advanced economies* in the *September 2006 World Economic Outlook*.[7]

We also include in our dataset a dummy variable that equals 1 if a country shares a *common language* with the U.S. (i.e., English) and another variable, *distance*, which measures the geographical *distance* (in miles) between a country's capital and Washington DC. Portes and Rey (2005) find in gravity equations that *distance* serves as an adequate proxy for information. Hence, we could potentially expect *distance* to be positively related to *home bias*. Moreover, *common language* and *distance* may affect *attention* since it proxies for cultural differences between the U.S. and other countries in our sample. The effect of cultural differences could potentially be either positive or negative. A positive effect suggests that agents use search engines to learn about more distant cultures, whereas a negative effect reveals that Americans are more interested in similar cultures.

Finally, Bohn and Tesar (1996) provide empirical evidence of U.S. investors' return chasing behavior. American investors tend to move into foreign markets where returns are expected to be high. The authors predict excess returns by regressing actual excess returns on lagged prediction variables. We include in our dataset one of the prediction variables: the lagged *stock market return* for each country in U.S. dollars. We collect daily data of the main local stock index and exchange rates from *Bloomberg News Service* and calculate the accumulated *stock market return* in 2005, the year prior to our *home bias* data.

Table 2 illustrates the summary statistics, and Table 3 provides the correlation matrix of the variables described in this section.

[**Insert Tables 2 and 3 about here**]

---

[7]The list consists of 29 countries: Australia, Austria, Belgium, Canada, Cyprus, Denmark, Finland, France, Germany, Greece, Hong Kong SAR, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, Taiwan Province of China, United Kingdom, United States. Note that Cyprus, Luxembourg, Taiwan Province of China and the United States are not included in our sample.

5   Methodology

We use an instrumental variable approach to disentangle the effects of attention allocation on home bias and vice versa. Specifically, we estimate a system of two cross-sectional equations using *home bias* and *attention* as the endogenous variables:

$$home\ bias_i\ =\ a_0 + a_1\ attention_i + a_2\ market\ sophistication_i + a_3\ enforce\ a\ contract_i + \varepsilon_i^{hb}$$

$$attention_i\ =\ b_0 + b_1\ home\ bias_i + b_2\ cultural\ sites_i + b_3\ female\ models_i + b_4\ internet\ users_i + \varepsilon_i^{att}$$

We use three instruments for *attention* allocation in the *home bias* equation. Internet users look for different pieces of information when using search tools. We focus on those that are unrelated to financial investments and use sensible proxies for them. A number of studies in computer science are dedicated to the topical classification of Internet user search queries.[8] Associating an often ambiguous and short search query to the right topic is challenging, yet essential for effectiveness and efficiency of Internet search engines. Pass et al (2006) provide a topical breakdown of search queries from the AOL web search service into 17 specific categories plus one additional residual group ("Other"). This categorization of Internet user search topics provides an excellent starting point. We eliminate four out of 17 categories which are potentially related to asset holdings ("Research", "Business", "News", and "Finance") and examine the remainder for potential instruments. We focus on two of the Top 5 most popular search query topics, "Places" and "Porn", which rank fifth and third, respectively.

Our first candidate for an instrumental variable is an indicator of a country's popularity in search queries for "Places": the number of World Heritage *cultural sites*. The Temple of Angkor Wat, the Taj Mahal, and the Pyramid Fields from Giza to Dahshur all attract attention to Cambodia, India, and Egypt, respectively, not because they are associated with financial investment opportunities, but due to their outstanding cultural value to humanity. Hence, it is natural to expect that the more *cultural sites* within a country the more *attention*

---

[8]See Beitzel, Jensen, Frieder, Grossman, Lewis, Chowdhury and Kolcz (2005a), Beitzel, Jensen, Frieder, Lewis, Chowdhury and Kolcz (2005b), He, Yan, Ma and Chen (2007), Joachims (2002), and Lewandowski (2006).

the country will attract.

The second candidate is related to "Porn": a ranking based on how popular a country's *female models* are amongst American citizens. The more popular a country's *female models*, the more *attention* the country will receive. Although this argument is sensible, we must consider the possibility of reverse causality. If a given country is more recognizable to Americans than others, its *female models* may be in demand and thus more likely to become famous world-wide. In this alternative case, the popularity of *female models* in a specific country would merely be a reflection of the extra attention the country receives because of familiarity effects generated by past common history.[9] In other words, stronger cultural proximity may be the main factor behind American society's demand for a country's cuisine, music, movies, art, models, consumer products and financial investments. If this alternative story is true, the inclusion of explanatory variables such as common language or geographical distance – the usual proxy for cultural proximity used in gravity models – in our regressions would have two important effects: *i*) *female models* would have no significance in explaining *attention* allocation, and *ii*) *attention* allocation would have no significance in explaining *home bias*. However, as we will show in our robustness checks, this is not the case. Empirical evidence does not seem to support this alternative story.

For our third potential instrument, we move away from search query topics and towards online content creation. Internet users create content online when they post their thoughts, photos, videos, music, or artwork on a personal Web site, Web blog or Web diary. It is natural to expect that countries with a larger share of Internet users engaged in content creation would receive more attention. Unfortunately, there is no data available on content creation intensity by country, which forces us to rely on a proxy. According to Lenhart et al (2004), approximately half of the adult American Internet users have contributed material to the online world. If we assume that this is also the case for other countries, then the share of Internet users within a country's population could proxy for online content creation intensity. Hence, we use the number of *Internet users* (per 100) as our third candidate for an

---

[9]Mondria and Wu (2010) show that familiarity of domestic agents with foreign countries, proxied by the number of international departures from the home country for any purpose other than a remunerated activity in the country visited, is an important determinant of home bias.

instrumental variable.

With regards to the instruments for *home bias*, it is natural to expect investors to shy away from markets with higher implicit costs, whether they are generated by lack of financial sophistication or low quality financial regulation. Moreover, there is no particular reason to expect implicit financial costs to affect *attention* through channels other than asset holdings. Therefore, we consider measures of implicit financial costs as instrumental variables candidates for *home bias* in the *attention* equation. The first candidate is a qualitative measure of each country's financial *market sophistication* level. The second candidate is the number of procedures required to *enforce a contract* in each country. The higher the number of procedures in a given country, the lower the quality of the regulatory framework.

6   Empirical Evidence

In this section, we present empirical evidence for the two-way causality between home bias and attention allocation. First, we estimate our baseline model using two different estimation procedures: equation-by-equation 2SLS and simultaneous equation 3SLS.[10] Then, we run robustness checks by either considering different subsamples or by testing the significance of additional variables that control for size, economic development, gravity, and return chasing.

6.1   Baseline Model

Our full sample baseline model estimation results are shown in Table 4. The table presents two pairs of equations, each estimated with a different methodology. Equations (4.1) and (4.2) were individually estimated by 2SLS,[11] while the system of equations (4.3) was jointly estimated by 3SLS. The most noteworthy feature of our results is that regardless of the estimation procedure, there is evidence of a significant two-way causality between *attention* and *home bias* as predicted by our model. Moreover, all models have a good in-sample fit. The 2SLS procedure yields an R^2 of 65.9% for the *attention* equation and 48.2% for the *home bias*

---

[10] There is a trade-off between both methods. While the 3SLS estimator is more efficient, the 2SLS estimator is more robust. The former procedure requires the system to be perfectly specified. A wrong instrument in one equation will bias the coefficients of other equations from the system.

[11] See Appendix for the 2SLS first stage regression estimation outputs.

equation. The 3SLS methodology produces slightly smaller R²'s, of 60.2% for *attention* and 47.5% for *home bias*.[12]

<p style="text-align:center">[Insert Table 4 about here]</p>

All estimated coefficients in the *attention* equations are significant at the 1% level. The number of procedures required to *enforce a contract* and the level of financial *market sophistication* are used as instruments for *home bias*. The results reveal that agents do allocate more *attention* to countries in which *home bias* is smaller: a one percentage point decrease in the U.S. *home bias* towards a particular country is associated with either a 9.98% or 11.22% increase in the number of that country's click-throughs if estimated by 2SLS or 3SLS, respectively.[13] We can also see that *attention* is affected by the number of World Heritage *cultural sites*, the popularity of a country's *female models*, and the number of *internet users* in a given country. A 1% increase in *cultural sites* increases *attention* by either 0.44% or 0.55%, a 1% increase in *female models* increases *attention* by either 0.25% or 0.20%, and a 1% increase in *internet users* increases *attention* by 0.62% or 0.47%. All of the previous results depend on whether the coefficient is estimated by 2SLS or 3SLS, respectively.

With respect to the *home bias* equations, we use *cultural sites*, *female models* and *internet users* as instrumental variables for the *attention* allocated to each country. The results illustrate that U.S. investors have less *home bias* towards countries in which they allocate more *attention*: a 10% increase in the number of click-throughs received by a country decreases *home bias* by 0.27 or 0.30 percentage points, depending on whether the coefficient is estimated by 2SLS or 3SLS, respectively.[14] Both coefficients are significant at the 5% level. There is also less *home bias* from U.S. investors in countries with more sophisticated financial markets: a 1-point increase in *market sophistication* decreases *home bias* by 4.8 (2SLS estimator) or 3.5 (3SLS estimator) percentage points, with the coefficient being significant at the 5% level. The

---

[12]In-sample fits are better if *home bias* is not in natural logs. However, all results are qualitatively the same regardless of whether natural logs are taken or not.

[13]Note that *home bias* is calculated as a percentage and that it is not included in the regressions in natural logs. Therefore, when we state that US *home bias* towards a particular country, i.e. Spain, decreases by one percentage point, we mean that Spain's *home bias* decreases from 83.8% to 82.8%.

[14]Once again, *home bias* is calculated as a percentage and it is not included in the regressions in natural logs. Hence, coefficients refer to the estimated absolute percentage point change in *home bias*.

number of procedures required to *enforce a contract* is only significant at the 5% level in the 3SLS estimation: for each additional procedure that is necessary to *enforce a contract*, there is a 0.4 percentage point increase in the amount of *home bias*.

Overidenfication tests do not reject the validity of the chosen instruments in any procedure. In other words, we cannot reject the orthogonality conditions assumed when estimating the system either by 2SLS or 3SLS. The Hansen-Sargan statistics in result of the 2SLS methodology are 1.051 for the *attention* equation and 1.738 for the *home bias* equation with p-values of 30.5% and 41.9%, respectively. Furthermore, the Hansen-Sargan statistic obtained by the 3SLS estimation is 2.545 with a p-value of 46.7%.

[Insert Figure 2 about here]

One exercise helps us quantify the importance of attention allocation in explaining home bias. We calculate counterfactual values of U.S. investors' home bias towards all countries in our sample under a "no home attention bias" assumption. More specifically, we assume that all countries in our sample receive the exact same amount of attention that U.S. investors allocate to their home country.[15] Under this assumption and using the coefficient estimated by 3SLS, average U.S. home bias in our sample falls from 85.2% to 57.3%. There are also large reductions in minimum and maximum home bias levels in our sample: maximum home bias falls from 100% to 74.4% and minimum home bias from 26.5% to 5.9%. Figure 2 highlights the differences between the inverse cumulative distributions of home bias in the actual data and that in our simulation.

6.2 Robustness to Alternative Country Samples

We examine the robustness of our results with alternative country samples. As noted by Ahearne et al (2004), U.S. holdings of foreign equities are unevenly distributed across countries. U.S. investors concentrate a significant share of their foreign holdings on a handful of countries and hold virtually no equities in a large number of countries. The first pair of equations in Table 5 contain the estimation output of the baseline model with a subsample that excludes the

---

[15]In the AOL dataset, the *attention* level users allocated towards the US is 12,467,251.

Top 5 countries, which at the end of 2006 constituted 48.17% of U.S. foreign equity holdings. The Top 5 countries are the United Kingdom, Japan, France, Canada, and Switzerland. The second pair of equations in Table 5 present the regression results of our baseline model after eliminating fifteen countries within our sample towards which U.S. home bias is essentially 100%.

[Insert Table 5 about here]

The results confirm that the significant two-way causality between *attention* and *home bias* is not driven by either extremes of the U.S. holdings spectrum. All estimated coefficients reported in Table 5 are similar in magnitude and significance relative to those obtained with the full sample. Moreover, the Hansen-Sargan statistics still fails to reject the validity of the chosen instrument and the in-sample fits remain high.

6.3 Robustness to Additional Control Variables

We also check the robustness of our results to country size, development level, gravity, and return chasing. Omission of a control variable that simultaneously affects *attention* and *home bias* would generate correlation between both variables that could be mistakenly interpreted as causality. However, we find that evidence of a significant two-way causality between our dependent variables is not a consequence of model misspecification.

The first pair of equations in Table 6 show that our results are not driven by country size. In the attention equation, *population* and *land area* have coefficients that are not statistically significant. Moreover, country size has no effect on *home bias* either, as both variables also render insignificant coefficients. This result is not surprising, since *home bias* is essentially a linear transformation of U.S. holdings of foreign equities normalized by the size of a country's stock market. The second pair of equations in Table 6 suggest that a country's economic development level may affect *home bias* but not *attention.* In the *attention* equation, neither the *advanced economies* dummy nor *GDP per capita* variable are statistically significant. With respect to *home bias*, *GDP per capita* is not significant while *advanced economies* is significant at the 5% significance level. The statistically significant coefficient on *advanced*

19

economies implies that U.S. investors have on average 10.1 percentage points less *home bias* against advanced economies.

[**Insert Table 6 about here**]

The first pair of equations in Table 7 reveal that coefficients associated to *common language* and *distance* are not statistically significant. The cross-sectional variation in *attention* is most likely already capturing informational asymmetries present in investment decisions. Moreover, the second pair of equations in Table 7 show no evidence of return chasing behavior in our dataset as the coefficient associated to past *stock market return* is not significant.

[**Insert Table 7 about here**]

Once again, we would like to highlight that our results are robust to additional controls for size, economic development, gravity and return chasing. That is, all estimated coefficients presented in Tables 6 and 7 which are associated with variables from our baseline model have similar values and significance relative to those presented in Table 4. Moreover, the orthogonality condition assumed is not rejected by Hansen-Sargan statistics in any regression.

7   Conclusion

On August 3, 2006, AOL released a dataset which includes over 21 million web search queries from 657,426 of its customers. The data includes all searches and results from users for a three month period as well as a record regarding whether or not the user clicked on a search result. Using the AOL dataset, we constructed a measure of cross-country attention allocation, given by the number of times users clicked through a search result provided by a particular country.

We combine the click-through series with 2006 home bias data in the U.S.. Using an instrumental variables approach, we find empirical evidence of two-way causality between attention allocation and home bias. We use the number of World Heritage cultural sites, female models popularity rank, and percentage of internet users in each country as instrumental variables for attention allocation. Home bias is instrumented by the number of procedures

required to enforce a contract and the level of financial market sophistication in a country. We observe that our results are not driven by the handful of countries which receive almost half of U.S. investment, nor by the large number of countries in which virtually no assets are held by U.S. investors. Our results are also robust to the inclusion of additional variables that control for size, economic development, gravity, and return chasing. We calculate counterfactual values of U.S. investors' home bias towards all countries in our sample under a "no home attention bias" assumption. More specifically, we assume that all countries in our sample receive the exact same amount of attention that U.S. investors allocate to their home country. Under this assumption and using the coefficient estimated by 3SLS, average U.S. home bias in our sample falls from 85.2% to 57.3%.

To our knowledge, this is the first paper to measure attention allocation using internet search engine tools data. This approach is intuitive for at least two reasons. First, the World Wide Web is becoming a predominant information media. Second, search engines are the most popular tool used in helping users find reliable information on the Web since it minimizes the time required and the amount of information which must be consulted. In light of the increasing popularity of the World Wide Web and search engines as well as our positive results in this paper, we encourage the further use of search query data in the empirical analysis of attention allocation theories.

References

[1] Ahearne, A., Griever, W., Warnock, F., 2004. Information Costs and Home Bias: An Analysis of U.S. Holdings of Foreign Equities. Journal of International Economics 62, 313-336.

[2] Bacchetta, P., Van Wincoop, E., 2009. Infrequent Portfolio Decisions: A Solution to the Forward Discount Puzzle. American Economic Review, forthcoming.

[3] Barron, J., Ni, J., 2008 .Endogenous Asymmetric Information and International Equity Home Bias: The Effects of Portfolio Size and Information Costs. Journal of International Money and Finance 27, 617–635.

[4] Beitzel, S., Jensen, E., Frieder, O., Grossman, D., Lewis, D., Chowdhury, A., Kolcz, A., 2005. Automatic Web Query Classification Using Labeled and Unlabeled Training Data.

Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, 581-582.

[5] Beitzel, S., Jensen, E., Frieder, O., Lewis, D., Chowdhury, A., Kolcz, A., 2005. Improving Automatic Query Classification Via Semi-supervised Learning. Proceedings of the Fifth IEEE International Conference on Data Mining, 42-49.

[6] Bekaert, G., 1995. Market Integration and Investment Barriers in Emerging Equity Markets. World Bank Economic Review 9, 75–107.

[7] Bohn, H., Tesar, L., 1996. U.S. Equity Investment in Foreign Markets: Portfolio Rebalancing or Return Chasing?. American Economic Review 86, 77-81.

[8] Brennan, M., Cao, H., 1997. International portfolio flows. Journal of Finance 52, 1851-1880.

[9] Chan, K., Covrig, V., Ng, L., 2005. What Determines the Domestic Bias and Foreign Bias?. Journal of Finance 60, 1495-1534.

[10] Coval, J., Moskowitz, T., 2001. The Geography of Investment: Informed Trading and Asset Prices. Journal of Political Economy 109, 811-841.

[11] Dahlquist, M., Pinkowitz, L., Stulz, R., Williamson R., 2003. Corporate Governance and the Home Bias. Journal of Financial and Quantitative Analysis 38, 87-110.

[12] Dahlquist, M., Robertsson, G., 2001. Direct Foreign Ownership, Institutional Investors, and Firm Characteristics. Journal of Financial Economics 59, 413-440.

[13] Falkinger, J., 2007. Attention Economies. Journal of Economic Theory 133, 266-294.

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

[15] French, K., Poterba, J., 1991. Investor diversification and international equity markets. American Economic Review 81, 222–226.

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

[17] Gehrig, T., 1993. An Information Based Explanation of the Domestic Bias in International Equity Investment. Scandinavian Journal of Economics 95, 97–109.

[18] Hatchondo, J., 2008. Asymmetric Information and the Lack of International Portfolio Diversification. International Economic Review 49, 1297-1330.

[19] He, X., Yan, J., Ma, J., Liu, N., Chen, Z., 2007. Query Topic Detection for Reformulation. Proceedings of the 16th international conference on World Wide Web, 1187-1188.

[20] Joachims, T., 2002. Optimizing Search Engines Using Clickthrough Data. Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, 133-142.

[21] Kang, J., Stulz, R., 1997. Why Is There a Home Bias? An Analysis of Foreign Portfolio Equity Ownership in Japan. Journal of Financial Economics 46, 3-28.

[22] Kraay, A., Loayza, N., Serven, L., Ventura, J. 2005. Country Portfolios. Journal of the European Economic Association 3, 914–945.

[23] Lewandowski, D., 2006. Query Types and Search Topics of German Web Search Engine Users. Information Services & Use 26, 261-269.

[24] Lenhart, A., Horrigan, J., Fallows, D., 2004. Content Creation Online. Washington, DC: Pew Internet & American Life Project.

[25] Luo, Y., 2008. Consumption Dynamics under Information Processing Constraints. Review of Economic Dynamics 11, 366-385.

[26] Mackoviak, B., Wiederholt, M., 2009a. Optimal Sticky Prices under Rational Inattention. American Economic Review 99, 769-803.

[27] Mackowiak, B., Wiederholt, M., 2009b. Business Cycle Dynamics under Rational Inattention. Working paper.

[28] Malloy, C., 2005. The Geography of Equity Analysis. Journal of Finance 60, 719-755.

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

[30] Mondria, J., 2010. Portfolio Choice, Attention Allocation, and Price Comovement. Journal of Economic Theory, forthcoming.

[31] Mondria, J., Wu, T., 2010. The Puzzling Evolution of the Home Bias, Information Processing and Financial Openness. Journal of Economic Dynamics and Control 34, 875-896.

[32] Pass, G., Chowdhury, A., Torgeson, C., 2006. A Picture of Search. Proceedings of the 1st international conference on Scalable information systems.

[33] Peng, L., Xiong, W., 2006. Investor Attention, Overconfidence and Category Learning. Journal of Financial Economics 80, 563-602.

[34] Portes, R., Rey, H., 2005. The Determinants of Cross Border Equity Flows. Journal of International Economics 65, 269-296.

[35] Sims, C., 2006. Rational Inattention: A Research Agenda. American Economic Review 96, 158-163.

[36] Tesar, L., Werner, I., 1995. Home bias and high turnover. Journal of International Money and Finance 14, 467–492.

[37] The Global Competitiveness Report, 2008. World Economic Forum.

[38] The Travel & Tourism Competitiveness Report, 2008. World Economic Forum.

[39] Tutino, A., 2009. The Rigidity of Choice . Lifecycle Savings with Information-Processing Limits. Working paper.

[40] Van Nieuwerburgh, S., Veldkamp, L., 2009. Information Immobility and the Home Bias Puzzle. Journal of Finance 64, 1187-1215.

[41] Van Nieuwerburgh, S., Veldkamp, L., 2010. Information Acquisition and Portfolio Under-Diversification. Review of Economic Studies, forthcoming.

[42] Zhou, C., 1998. Dynamic portfolio choice and asset pricing with differential information. Journal of Economic Dynamics and Control 22, 1027–1051.

A   Appendix

Table 8 presents the estimation output of the 2SLS first stage *attention* and *home bias* regressions. There are a couple of features in these results that are worth commenting. First, we can note that *cultural sites*, *female models* and *internet users* are highly significant in the *attention* equation. With respect to instruments for *home bias*, only *market sophistication* yields a significant coefficient. The coefficient associated to *enforce a contract*, with a t-statistic of 1.35 and a p-value of 0.182, is not significant. However, we can also note that *enforce a contract* is only significant in the 3SLS procedures.

The results also show that both *enforce a contract* and *market sophistication* are significant in the *attention* equation. Most likely, both variables are capturing the strong effect of *home bias* on *attention* allocation. The first stage coefficients of both instruments is similar in magnitude to their 2SLS second stage coefficients in the *home bias* equation multiplied by the 2SLS *home bias* coefficient in the *attention* equation. This is probably the same reason why *cultural sites* is significant in the first stage 2SLS *home bias* equation.

Figure 1: AOL Attention Allocation and International Investment
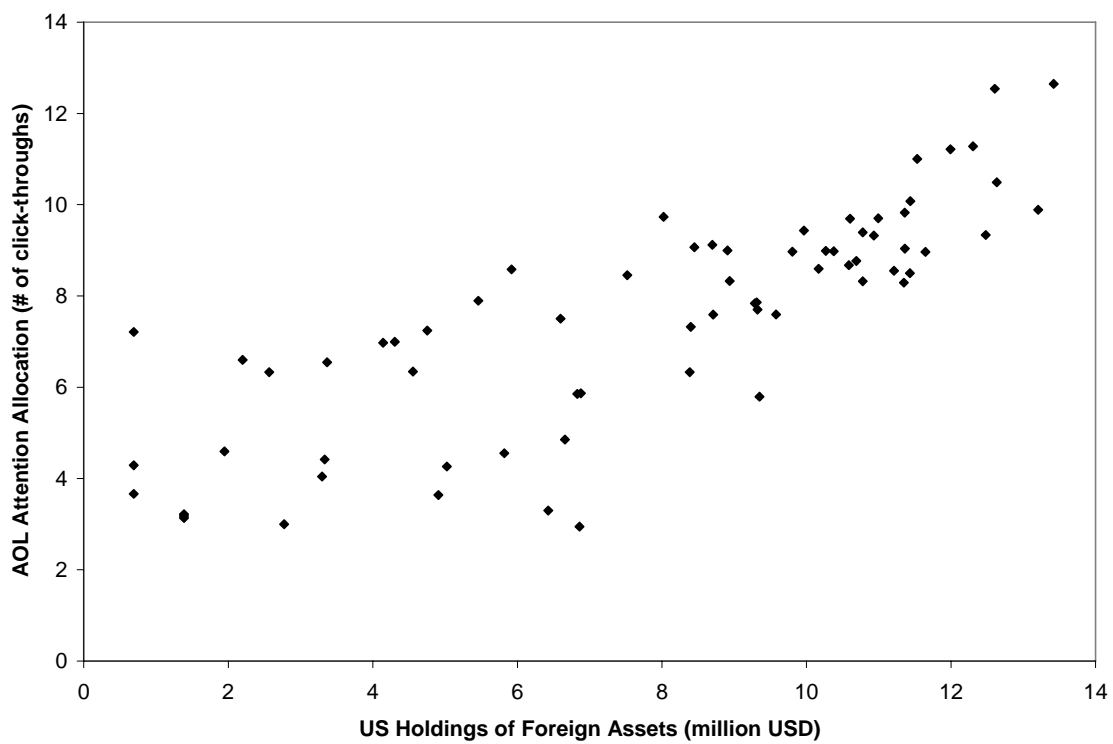
Note: both variables in natural logs.

Figure 2: Inverse Cumulative Distribution of Actual and Simulated Home Bias
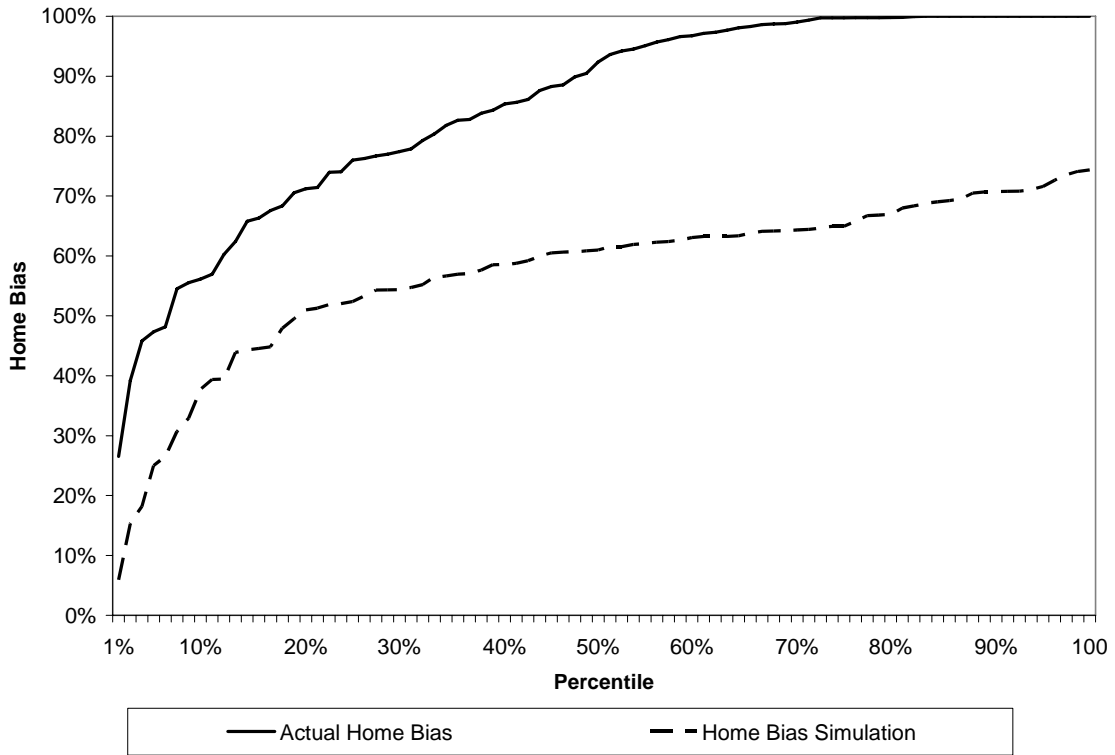
Note: countries are ordered from smallest to largest home bias in the horizontal axis. Simulated home bias is calculated assuming that all countries in our sample receive the exact same amount of attention that US investors allocate to their home country and using the coefficient estimated by 3SLS in our baseline model (Table 4).

Table 1: Attention Allocation, Female Model Popularity Rank and Home Bias Statistics

Country Name	Attention Allocation	Female Model	Home Bias	Country Name	Attention Allocation	Female Model	Home Bias
Argentina	4,690	0	94.2%	Latvia	560	0	98.8%
Armenia	43	0	100.0%	Lithuania	734	0	99.8%
Australia	60,103	228	76.7%	Macedonia, FYR	74	0	100.0%
Austria	7,851	0	76.3%	Malaysia	2,531	0	88.5%
Bangladesh	25	0	99.7%	Mauritius	27	0	56.9%
Belgium	8,007	0	81.8%	Mexico	3,987	208	39.2%
Bolivia	50	0	100.0%	Mongolia	34	0	100.0%
Botswana	23	0	99.7%	Morocco	95	0	98.3%
Brazil	4,918	412	67.6%	Namibia	73	0	99.1%
Bulgaria	569	0	97.7%	Nepal	36	0	100.0%
Canada	279,367	1,197	56.1%	Netherlands	74,022	0	48.2%
Chile	1,514	0	93.6%	New Zealand	8,677	71	74.0%
China	5,182	37	92.4%	Norway	7,923	31	71.4%
Colombia	1,815	190	96.7%	Pakistan	128	0	95.7%
Costa Rica	1,358	0	99.7%	Paraguay	64	0	100.0%
Croatia	1,090	0	99.4%	Peru	348	0	96.1%
Czech Republic	16,901	232	84.3%	Philippines	1,978	0	77.9%
Denmark	12,505	0	77.0%	Poland	8,073	40	87.6%
Ecuador	343	0	100.0%	Portugal	9,157	0	85.6%
Egypt, Arab Rep.	562	0	88.3%	Romania	5,335	0	97.2%
El Salvador	184	0	100.0%	Russian Federation	16,241	194	90.5%
Estonia	1,070	0	97.4%	Singapore	6,422	0	60.2%
Finland	11,185	0	47.3%	Slovak Republic	2,243	0	100.0%
France	36,005	81	68.4%	Slovenia	1,398	33	98.1%
Georgia	57	0	89.9%	South Africa	5,860	298	86.1%
Germany	79,333	197	66.3%	Spain	18,471	178	83.8%
Greece	1,986	0	82.6%	Sri Lanka	71	0	95.1%
Hong Kong, China	8,415	0	76.0%	Sweden	16,323	83	74.1%
Hungary	4,125	0	54.5%	Switzerland	11,341	0	45.8%
Iceland	695	0	99.8%	Tanzania	40	0	100.0%
India	4,122	54	85.3%	Thailand	2,593	0	80.4%
Indonesia	328	0	79.3%	Trinidad and Tobago	71	0	100.0%
Ireland	12,036	0	26.5%	Tunisia	24	0	99.8%
Israel	5,401	200	62.4%	Turkey	2,210	0	82.8%
Italy	23,782	95	77.4%	Uganda	11	0	100.0%
Jamaica	39	0	100.0%	Ukraine	2,683	54	98.6%
Japan	19,657	26	71.2%	United Kingdom	311,326	1,195	55.5%
Jordan	83	0	99.8%	Uruguay	172	0	100.0%
Kazakhstan	19	0	94.5%	Venezuela, RB	353	0	70.5%
Kenya	99	0	99.8%	Vietnam	299	0	100.0%
Korea, Rep.	7,830	22	65.8%	Zambia	20	0	96.6%
Kyrgyz Republic	8	0	100.0%	Zimbabwe	38	0	98.7%

Table 2: Summary Statistics

	Number of Observations	Mean	Median	Standard Deviation	Minimum	Maximum
attention	84	13,636	1,665	46,511	8	311,326
home bias	84	85.2%	93.0%	17.4%	26.5%	100.0%
cultural sites	84	6.8	0	8.5	0	40
female models	84	64	4	195	0	1,197
internet users	84	32.1	24	25.1	0.3	85.9
market sophistication	84	4.7	5.0	1.2	2.6	6.8
enforce a contract	84	35	36	6	20	50
population	84	60,900,000	12,449,078	187,000,000	303,782	1,310,000,000
land area	84	1,077,422	291,780	2,621,963	699	17,100,000
GDP per capita	83	14,480	5,923	16,761	327	72,325
advanced economies	84	0.30	0	0.46	0	1
distance	84	5,461	4,905	2,230	456	10,187
stock market return	68	26.2%	18.7%	36.2%	-39.2%	194.5%

	attention	*home bias*	*cultural sites*	*female models*	*internet*	*market sophistication*	*enforce a contract*	*population*	*land area*	*GDP per capita*	*advanced economies*	*distance*	*stock market return*
attention	100%												
home bias	-36%	100%											
cultural sites	26%	-21%	100%										
female models	91%	-31%	27%	100%									
internet	36%	-54%	15%	22%	100%								
market sophistication	36%	-61%	14%	32%	70%	100%							
enforce a contract	-15%	37%	17%	-2%	-52%	-44%	100%						
population	-3%	6%	39%	1%	-25%	-12%	26%	100%					
land area	24%	-2%	19%	39%	-11%	-4%	17%	41%	100%				
GDP per capita	38%	-65%	23%	24%	86%	73%	-51%	-20%	-8%	100%			
advanced economies	38%	-62%	32%	25%	77%	71%	-44%	-16%	-9%	88%	100%		
distance	-28%	16%	-27%	-29%	-29%	-9%	0%	18%	-3%	-30%	-15%	100%	
stock market return	-10%	19%	-12%	0%	-26%	-28%	26%	-3%	18%	-25%	-25%	-2%	100%

Table 4: Baseline Model Estimation Output

	(4.1)	(4.2)	(4.3)	
Estimation method:	2SLS	2SLS	3SLS	
Variables:	*attention*	*home bias*	*attention*	*home bias*
attention	-	-0.027***	-	-0.030***
	-	(0.010)	-	(0.009)
home bias	-9.975***	-	-11.22***	-
	(2.232)	-	(2.149)	-
cultural sites	0.437**	-	0.549***	-
	(0.207)	-	(0.178)	-
female models	0.246***	-	0.200***	-
	(0.093)	-	(0.061)	-
internet users	0.620***	-	0.467***	-
	(0.153)	-	(0.139)	-
enforce a contract	-	0.002	-	0.004**
	-	(0.003)	-	(0.002)
market sophistication	-	-0.048**	-	-0.035**
	-	(0.020)	-	(0.016)
Overidentification test				
Hansen-Sargan statistic	1.051	1.738	2.545	
	(0.305)	(0.419)	(0.467)	
Number of observations	84	84	84	
R^2	65.9%	48.2%	60.2%	47.5%

Note: Equations (4.1) and (4.2) were estimated individually by two stages least squares with White's robust standard errors. The system of equations (4.3) was jointly estimated 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. The following variables are in natural logs: *attention, internet users,* (one plus) *cultural sites* and (one plus) *female models.*

Table 5: Baseline Model with Alternative Country Samples

Variables:	(5.1)		(5.2)	
Estimation method:	**3SLS**		**3SLS**	
Variables:	*attention*	*home bias*	*attention*	*home bias*
attention	-	-0.030***	-	-0.029***
	-	(0.010)	-	(0.011)
home bias	-12.41***	-	-9.832***	-
	(2.448)	-	(2.272)	-
cultural sites	0.559***	-	0.534***	-
	(0.200)	-	(0.185)	-
female models	0.165***	-	0.224***	-
	(0.063)	-	(0.065)	-
internet users	0.433***	-	0.552***	-
	(0.140)	-	(0.172)	-
enforce a contract	-	0.004**	-	0.005*
	-	(0.002)	-	(0.002)
market sophistication	-	-0.028*	-	-0.040**
	-	(0.016)	-	(0.019)

Overidentification test

Hansen-Sargan statistic	2.212		2.691	
	(0.530)		(0.442)	
Sample	excludes Top 5 countries which attract the most US equity investment		excludes countries with less than $500,000 of US equity investment	
Number of observations	79		69	
R^2	46.5%	40.8%	57.1%	41.4%

Note: Each pair of 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. The following variables are in natural logs: *attention, internet users,* (one plus) *cultural sites* and (one plus) *female models.*

Table 6: Robustness Checks – Country Size and Development Level

Estimation method:	(6.1) 3SLS		(6.2) 3SLS	
Variables:	*attention*	*home bias*	*attention*	*home bias*
attention	-	-0.026**	-	-0.025**
	-	(0.012)	-	(0.011)
home bias	-10.83***	-	-12.55**	-
	(2.569)	-	(5.054)	-
cultural sites	0.314*	-	0.538***	-
	(0.165)	-	(0.200)	-
female models	0.160***	-	0.200***	-
	(0.062)	-	(0.068)	-
internet users	0.699***	-	0.402**	-
	(0.196)	-	(0.180)	-
enforce a contract	-	0.005**	-	0.003*
	-	(0.002)	-	(0.002)
market sophistication	-	-0.037**	-	-0.025
	-	(0.019)	-	(0.018)
population	0.161	-0.019	-	-
	(0.158)	(0.013)	-	-
land area	0.084	0.010	-	-
	(0.125)	(0.009)	-	-
advanced economies	-	-	-0.973	-0.101**
	-	-	(0.917)	(0.049)
GDP per capita	-	-	0.185	0.009
	-	-	(0.323)	(0.020)

Overidentification test

Hansen-Sargan statistic	1.831		2.256	
	(0.608)		(0.521)	
Number of observations	84		83	
R^2	63.6%	48.9%	53.8%	50.0%

Note: Each pair of 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. The following variables are in natural logs: *attention, internet users, population, land area,GDP per capita,* (one plus) *cultural sites* and (one plus) *female models.*

Table 7: Robustness Checks – Gravity and Return Chasing

	(7.1)		(7.2)	
Estimation method:	3SLS		3SLS	
Variables:	*attention*	*home bias*	*attention*	*home bias*
attention	-	-0.034***	-	-0.027**
	-	(0.010)	-	(0.012)
home bias	-12.99***	-	-8.014***	-
	(2.730)	-	(2.203)	-
cultural sites	0.431*	-	0.478***	-
	(0.230)	-	(0.172)	-
female models	0.182***	-	0.243***	-
	(0.062)	-	(0.065)	-
internet users	0.360**	-	0.656***	-
	(0.159)	-	(0.180)	-
enforce a contract	-	0.003*	-	0.005*
	-	(0.002)	-	(0.003)
market sophistication	-	-0.027	-	-0.050**
	-	(0.017)	-	(0.020)
common language	-0.748	-0.045	-	-
	(0.520)	(0.035)	-	-
distance	0.083	0.004	-	-
	(0.387)	(0.029)	-	-
stock market return	-	-	-0.425	0.011
	-	-	(0.733)	(0.070)

Overidentification test

Hansen-Sargan statistic	2.363		1.935	
	(0.501)		(0.586)	
Number of observations	84		68	
R^2	52.5%	48.3%	62.4%	43.0%

Note: Each pair of 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. The following variables are in natural logs: *attention, internet users, distance,* (one plus) *cultural sites,* (one plus) *female models* and (one plus) *stock market return.*

Table 8: Baseline Model 2SLS First Stage Estimation Output

	(8.1)	**(8.2)**
Estimation method:	**OLS**	**OLS**
Variables:	*attention*	*home bias*
cultural sites	0.880***	-0.038**
	(0.173)	(0.016)
female models	0.296***	-0.005
	(0.059)	(0.009)
internet users	0.671***	-0.006
	(0.154)	(0.014)
enforce a contract	-0.084***	0.005
	(0.024)	(0.004)
market sophistication	0.544***	-0.068***
	(0.132)	(0.016)
Number of observations	84	84
R^2	81.8%	46.7%

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. The following variables are in natural logs: *attention, internet users,* (one plus) *cultural sites* and (one plus) *female models.*