Essays on Similarities Between Executives and Investors

Harvey Cheong

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Jarrad Harford, Chair

Edward Rice

Stephan Siegel

Program Authorized to Offer Degree:
Michael G. Foster School of Business

University of Washington

Abstract

Essays on Similarities Between Executives and Investors

Harvey Cheong

Chair of the Supervisory Committee:
Professor Jarrad Harford
Department of Finance and Business Economics

The first chapter of my dissertation examines the effect of similarity in personal characteristics between start-up executives and investor principals on whether the start-up obtains funding from the investor, as well as the success of the start-ups. I find that while similarities in educational background, ethnicity, and cultural values between the start-up executives and investor principals are positively related to the likelihood of the start-up obtaining funding, these same similarities are negatively related to the likelihood of the start-up being acquired or going public.

The second chapter of my dissertation examines whether social connections between hedge fund managers and portfolio firm executives provide a channel through which corporate information can be transmitted to financial markets. We examine hedge fund investments involved in mergers and acquisitions as a potential setting in which time-sensitive corporate information may flow to equity markets. We find that a net put (call) position held by the hedge fund in the acquirer prior to a merger can predict the merger announcement returns being negative (positive), and that the predictive power increases with the depth of the social connection.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGMENTS

I am grateful to the Chair of my dissertation committee, Jarrad Harford, for his guidance and support throughout the doctoral program; Edward Rice and Stephan Siegel for their guidance and advice; Joon Ho Kim and Florian Münkel as co-authors; my cohort mates in the finance and accounting departments Darren Bernard, Aaron Burt, Nicole Cade, John Hackney, and Braden Williams as well as my friends Yucheng Low and Srikrishna Sridhar for discussion, assistance, and support; and finally, my family for everything.

Chapter two is co-authored with Joon Ho Kim and Florian Münkel.

DEDICATION

to my family

Chapter 1

THE RELATION BETWEEN SHARED CHARACTERISTICS AND PARTNER SELECTION AND SUCCESS: EVIDENCE FROM START-UPS

Economic agents frequently form partnerships with other economic agents for mutual benefit. The success of the partnership is dependent on the partners, as this not only determines the skills and resources that will be available to the partnership, but also affects the co-operation between partners after the partnership has been formed. So, how do economic agents find and choose their partners? Additionally, how do the criteria used to choose a partner relate to the success of the partnership?

Shared characteristics might matter for matching between economic agents. Economic agents may use the informational advantages that they have in their own social networks to find suitable partners within these networks. These social networks can be based on similarities: for example, a job-seeker may be able to find a position through the alumni network at their affiliated educational institutions, while similarly, an investor may be able to find entrepreneurs within their own ethnic communities. The shared characteristics may reduce the cost of assessing the suitability of a potential partner, such that the economic agent may be better able to ascertain the quality of a partner or mitigate the information asymmetry between them. However, shared characteristics not only influence partnership matching, but can also affect how the partnership functions after it has been formed. For example, shared characteristics can also facilitate better co-operation between partners, and therefore improve the chances of success or earning a positive return. However, it is not clear that shared characteristics necessarily improve economic outcomes, since economic agents

can also choose to work with partners who share similar characteristics at the expense of economic performance (Gompers, Mukharlyamov, and Xuan, 2016; Ishii and Xuan, 2014). Additionally, these similarities may also lead to the formation of a shared corporate culture between the economic agents based on similar beliefs. These shared beliefs can result in consistency in business decision making (Van den Steen, 2010a), which in turn improve the motivation of and communication between the economic agents but also reduce the monitoring between economic agents and the level of information collected by the agents themselves (Van den Steen, 2010b).

In this paper, I examine how shared characteristics or similarities between economic agents relate to the matching process between partners and the performance of the partnership in the context of start-ups and start-up investors. I choose to use this setting for several reasons. First, social networks and referrals, which are related to similarity in characteristics, are anecdotally important in the matching process between start-ups and start-up investors (Fried and Hisrich, 1994). Second, shared characteristics between start-up and start-up investor can reduce the level of information asymmetry between the investor and the start-up regarding the true quality of the start-up (Gompers and Lerner, 2001; Cochrane, 2006). Third, start-ups choose investors not only on the basis of the amount of funding offered, but also on the provision of support from the investor post-investment (Hellman and Puri, 2000; 2002), which may be improved by the shared characteristics between the start-up and its investors. Fourth and finally, start-ups and investors communicate frequently throughout the life of the start-up, thereby providing ample opportunities for shared characteristics to affect the co-operation between the partners and hence the performance of the start-up investment.

Using data from Crunchbase, a proprietary database of start-ups and start-up investors maintained by TechCrunch (a technology industry and start-up news portal), I examine how shared characteristics between start-up entrepreneurs and their investor principals affect the odds of being matched, as well as the odds of start-up success. I use this data-set primarily due to the availability of biographical information of the investor principals and the start-up

entrepreneurs, such as educational background. I assemble a sample of 9,138 investments from 2004 to 2013, which are comprised of 4,685 unique start-ups and 1,167 investors based in the United States. For each investment, I compute binary indicators denoting a shared educational background between the start-up entrepreneurs and the investor principals involved in an investment. The shared educational background is based upon the following: first, the same educational institution; second, the same type of degree; and third, the same educational institution and type of degree.

Additionally, I exploit the names of the individuals included in the data to create other variables of similarity. I use the first name of each individual to classify their gender, and create a corresponding indicator measure that indicates if at least one start-up entrepreneur and one investor principal involved with the investment are female. Similarly, I use the family names to derive a distribution of the likely ethnic groups of each start-up entrepreneur and investor principal, and create an indicator variable that indicates if at least one start-up entrepreneur and one investor principal share the same ethnicity. I also use the distribution of the likely ethnic groups of the individuals to compute an average cultural dimension score for each start-up and investor based on the cultural dimensions of Hofstede (1980; 1997; and 2001). I then calculate the Euclidean distance between the average cultural dimension scores for each start-up and investor, and use that as a proxy for the cultural distance between the start-up and investor. In order to control for geography specific preferences or differences in investment across years or industry, I also gather information on the geographic location of the start-up and investor, start-up industry, and year of funding of the start-up.

I first examine how start-ups and start-up investors match with each other. Since rejected matches between start-up and start-up investors are not observable, I create counter-factual investors for each investment by identifying other investors that are similar to the start-up investor for that investment. I select these counter-factual investors based on the actual start-up investor's preferences on start-up industry, investment stage, and start-up location based on information obtained from Thomson Reuters VentureXpert. I show that start-ups and investors whose personnel share similarities in educational background, ethnicity, and

cultural dimensions are more likely to match and partner with each other. Additionally, I show that these results are robust to using randomised start-up investors as counter-factual observations for each investment, suggesting that the choice of partner based on similarities is a choice made on purpose.

Next, I investigate the odds of start-up success conditional on the fact that the start-up and investor have been matched together. I define start-up as having succeeded when it has been acquired or has gone public within five or seven calendar years of funding as a measure of start-up success. Contrary to my findings regarding the matching between start-ups and investors, I show that start-ups whose personnel share similarities with the start-up's investor principals have at worst reduced odds and at best unchanged odds of exiting the start-up stage within five or seven calendar years of the first funding round. I also implement a Heckman selection correction to control for selection bias, since the matching process may have resulted in a set of start-ups which are most likely to succeed. The prior results would therefore not be indicative of how similarities relate to start-up success. Despite correcting for this potential selection bias, the results are consistent with the interpretation that similarities negatively affect the odds of start-up success. Therefore, while shared characteristics are related to increased odds that a start-up and an investor are matched together, the very same shared characteristics are also related with a lower probability of the start-up exiting the start-up stage.

A possible interpretation of the results is that investors are less discerning when investing in start-ups whose personnel share similarities with the investor principals. This could be because the investor principals over-value the synergy gains between themselves and similar start-up personnel, or that the investor principals are over-confident in their ability to assess the true quality of the start-up when the start-up personnel are similar. Another interpretation is that investor principals may have a preference to invest in start-ups whose entrepreneurs are similar simply as a matter of taste, and are willing to accept reduced performance in order to satisfy this preference. Alternatively, in contrast to high quality start-ups that would have no shortage of willing investors, it could be that lower quality

start-ups rely on being matched with a similar investor. Similarities would therefore appear to have a negative effect or relation with start-up performance.

This study contributes to the extant literature in several ways. First, I provide evidence on how economic agents choose business partners, specifically, showing that investor principals are more likely to match with similar start-up entrepreneurs. This is consistent with the model of Van den Steen (2010a), which shows that economic agents prefer to work with others who have similar beliefs, as these others would make similar decisions or choices as the economic agents themselves. Second, I present evidence on the performance implications of these similarities. I show that the similarities between start-up entrepreneurs and investor principals are at best unrelated and at worst negatively related to the odds that the start-up succeeds, as measured by start-up exits within five or seven calendar years of being first funded. This suggests that overall the net effect of corporate culture homogeneity is positive, as outlined in Van den Steen (2010b). Further, these implications may also be applicable to other business relationships, such as between employer and employee or buyers and suppliers (Wuyt and Geyskens, 2005). Finally, I use a novel proprietary database, Crunchbase, and show that it can be used in applications relevant to the finance literature. I also provide information on the database, including information on the structure and details of the database, as well as how to obtain it for future use.

1.1 Prior literature and hypotheses

The tendency of individuals to bond together with others based on shared characteristics has been well documented (McPherson, Smith-Lovin, and Cook, 2001). These shared characteristics therefore lead to bonding between similar people, which in turn leads to information sharing between them. In the context of start-ups, anecdotal evidence suggests the importance of social networks and referrals in the matching process between start-ups and investors (Fried and Hisrich, 1994), suggesting that social networks expand the set of possible investors that a start-up may reach (and vice versa).

However, other than the expanded set of potential partners, start-up entrepreneurs and

investor principals may choose to match with each other on the basis of similarities for other reasons. First, individuals may have a preference for working with others who are similar to themselves, a tendency which has been observed in the formation of entrepreneurial teams (Ruef, Aldrich, and Carter, 2003). This preference may be due to (i) a desire to improve the social standing of the shared group or (ii) similarities providing a reduced cost of assessing the true quality of the partner. An investor may have soft information that originates from the similarities that provides the investor with the ability to better evaluate the start-up team (Berger, Miller, Petersen, Rajan, and Stein, 2005). Alternatively, the investor may be able to call upon his or her common contacts within the shared network for a second opinion on the quality of the start-up or start-up team (Glaeser, Laibson, Scheinkman, and Soutter, 2000). Second, entrepreneurs and investor principals may use their common social networks as a means of enforcing the trust between them. News of misconduct by either party can be spread more quickly within their social networks than without, thereby increasing the possible severity of the reputation loss due to the misconduct for the misbehaving party. These two reasons can motivate economic agents to rely on shared similarities and the resulting social network to find and select partners, such as trade credit providers (Fisman, 2003). My first hypothesis is therefore:

Hypothesis 1. The odds of matching between start-up and investor is increased when the start-up entrepreneurs and investor principals share similarities.

Once the match has been made, the natural next step is to consider how these similarities may relate to the outcomes for the partnership. Similarities between two individuals can improve performance due to better information sharing, communication, or decision making. Mutual fund managers earn higher returns on portfolio firms whose entrepreneurs or board members share an educational history with the mutual fund managers themselves (Cohen, Frazzini, and Malloy, 2008), while financial analysts make better stock recommendations for firms whose entrepreneurs or board members have a similar educational background as the analysts (Cohen, Frazzini, and Malloy, 2010). Further, these similarities could also improve communication and information-sharing between the investor and entrepreneurs at board

meetings after investment has been made.

However, the presence of these similarities could also be detrimental to the partnership. Mergers and acquisitions that reduce shareholder value are more likely when the target and acquirer are socially connected (Ishii and Xuan, 2014), while venture syndicates that share characteristics are more likely to under-perform (Gompers, Mukharlyamov, and Xuan, 2016). In this case, the partners may prioritise conformity or harmony such that they ignore alternative opportunities or do not critically evaluate the pros and cons of the agreed upon decision. Additionally, these individuals may prefer to work with similar people, and therefore be more willing to accept reduced performance in order to be able to work with similar people. Another possibility is that an investor may even choose to work with a lower quality start-up that he or she shares similarities with, but aim to or believe that he or she can improve it based on the synergies that their similarities may bring.

Similarly, Van der Steen (2010a) outlines a model that shows that organizations innately tend towards shared beliefs, as people prefer to work with others who are more likely to make the same business decisions as themselves. However, while the shared beliefs may have positive affects on business operations, such as increased motivation, coordination, and communication, these very same beliefs can lead to reduced monitoring, experimentation, and information collection (Van den Steen, 2010b). Since beliefs may be informed by characteristics, shared characteristics can indicate similarities in beliefs between individuals. The net relation between (i) the similarity of start-up entrepreneurs and investor principals and (ii) start-up success is therefore uncertain. However, Hegde and Tumlinson (2014) provide evidence that the odds of start-up exit increases when the start-up entrepreneurs and investor principals have similar ethnic backgrounds. My second hypothesis is therefore that the odds of start-up exit is increased in the presence of similarities between the start-up entrepreneurs and investor principals, but this will ultimately be an empirical question to be answered.

Hypothesis 2. The odds of start-up exit is increased when the start-up entrepreneurs and investor principals share similarities.

Now that the hypotheses have been established, the next issue is to narrow down the

specific definition for similarity. Given the highly relationship-based nature of start-ups and start-up investments, as well as the importance of start-up entrepreneurs and investor principals on start-up success (Hsu, 2007; Sørenson, 2007), the social networks could be based on several dimensions. For example, co-ethnicity between start-up entrepreneurs and investor principals have been shown to be a predictor of matching between start-ups and investors. Bengtsson and Hsu (2015) focus on several ethnicities or groups that they deem to be important subgroups that are active in the U.S. venture capital industry[1], and find that co-ethnicity in their selected sub-groups increases the odds that a venture capital firm invests in a start-up. However, they do not examine start-up outcomes in relation to co-ethnicity. However, Hegde and Tumlinson (2014) develop a model of the effect of co-ethnicity on the matching between start-ups and venture capital firms, and show that investors lower their investment quality thresholds when a co-ethnic link is present, and that both the odds of matching and start-up exit are increased in the presence of co-ethnicity.

I focus on three separate categories of similarity. The first is based on social networks, which encompasses educational and ethnic similarity of the start-up entrepreneurs and the investor principals. My choice of the former is based on the work of Cohen, Frazzini, and Malloy (2008; 2010), where educational similarities are related to significant performance implications, while the latter follows the work of Bengtsson and Hsu (2015) and Hegde and Tumlinson (2014), where co-ethnicity is shown to be positively related to the odds that a start-up is (i) funded and (ii) exits the market. The second is based on similarities in culture, which I base on the cultural dimensions associated with the likely ethnicities of the start-up entrepreneurs and investor principals. While cultural dimensions have been shown to affect firm risk culture (Pan, Siegel, and Wang, 2016a), firm acquisitions (Pan, Siegel, and Wang, 2016b), and firm profitability (Nguyen, Hagendorff, Eshraghi, 2016), cultural dimensions per se do not provide a measure of the similarity between the two parties. Instead, I define a measure of cultural distance, which has been used in examining the internationalisation

[1]The ethnicities are: Chinese, Indian, Japanese, Korean, Russian, Hispanic, and Vietnamese.

strategies of corporations (Kogut and Singh, 1988), as well as the home-bias and foreign-bias of institutional investor holdings (Karolyi, 2016). Finally, for the last category of similarity, I choose to examine gender. Gender could be viewed as a similarity based on both network and values, given the presence of gender-based advocacy groups active in the technology industry, such as Women in Technology. Raina (2017) shows that female-led start-ups are less likely to exit if the start-up is funded by an all-male venture capital firm, which suggests that the presence of a female on both the start-up entrepreneurs and investor principals is positively related to the odds of start-up exit. Gompers, Mukharlyamov, Weisburst, and Xuan (2014) also show that female venture capitalists under-perform their male counterparts, suggesting a difference in start-up performance along gender lines.

1.2 Data

I use information on start-up investments by financial organisations to examine the effect of shared characteristics between start-up entrepreneurs and investor principals on the matching between start-ups and investors and success of the start-up. Using data from Crunchbase, I build a data set of 9,138 investments with corresponding data on the associated personnel of each investment, and compute measures of educational, ethnic, and gender similarities, as well as cultural distance. I gather data from several other sources, which are discussed in detail in the following sections.

1.2.1 Crunchbase and VentureXpert

The main data used in this paper comes from Crunchbase, a proprietary database of start-ups and venture capitalists created in 2007 by TechCrunch, which is a publisher of technology industry and start-up news. While limited access to the up-to-date data through the website is free, obtaining the complete, up-to-date data in a form suitable for analysis requires an active paid subscription. However, Crunchbase provides a complete snapshot of the data from December 2013 free of charge for registered users. The December 2013 snapshot comprises of a set of eleven data files, which cover the following information: acquisitions; funding

rounds; funds; investments; IPOs; start-up milestones; personnel information; personnel education; office locations; relationships and affiliations; and miscellaneous data. These files are cross-linked via person or entity identifiers.

Crunchbase obtains its data from three primary sources. First, it extracts information from on-line news sources and SEC filings, such as financing rounds, acquisitions, and management changes. Second, Crunchbase obtains data through data sharing partnerships with various market players via the Crunchbase Venture Program, which provides funders with free access to the full database in return for regular data sharing and updates. For example, Crunchbase has a data sharing partnership in place with AngelList, an on-line portal for people interested in investing and working at start-ups. Third, Crunchbase crowdsources data from at least 2 million users, of which an estimated 50,000 are active contributors. In order to ensure accuracy and data integrity, Crunchbase employs data editors to edit and check the information submitted to Crunchbase.

The key advantage of Crunchbase that I exploit for this study is the availability of biographical information, such as educational history, that is generally unavailable in other start-up and venture capital databases. I use the biographical data to build various measures of similarity between the start-up entrepreneurs and investor principals for the analysis. Another advantage of Crunchbase lies in the difference in scope. For example, Thomson Reuters VentureXpert, another commonly used start-up and venture capital database, relies primarily on venture capital fund reports for its data. This therefore limits VentureXpert data to start-ups that have obtained funding from venture capital funds. In contrast, Crunchbase includes start-ups that have obtained financing from alternative means, such as self-funding, crowdfunding, or angel investors. Additionally, since data is automatically added from Internet sources, concerns regarding self-reporting are somewhat mitigated, compared to VentureXpert, which depends on capital fund reports, whose providers may have an incentive to only report positive news.

Unfortunately, due to the structure of the data, investment round dates contained in the Crunchbase data do not appear to be completely reliable. I therefore choose to focus on the

first funding round for all start-ups. This restriction also results in a sample where the true quality of the start-up is most uncertain, and hence results in a setting where information sharing and communication can most directly improve the quality of the outcome. On the other hand, however, this restriction takes away the possibility of considering alternative measures of success, such as continued funding rounds and increases in valuation between rounds. Additionally, in order to eliminate the possibility of cross-border effects, I choose to focus on investments that are based in the United States.

I identify start-up and start-up investor pairs based on the first funding round. Due to the paucity of educational data for angel investors in the Crunchbase data, I choose to focus only on start-ups that have obtained funding from financial institutions, such as banks and private equity funds, as the educational data for the associated personnel is usually more complete. I use start-up exits within five or seven calendar years of the first funding round as a measure of start-up success in order to standardise the time frame for success across all start-ups. The five and seven calendar year windows are based on the Crunchbase data: the mean and median time to exit in the Crunchbase data is slightly above four years, while the 90-th percentile of the time to exit is about seven years.

Table 1.1 Panel A describes the distribution of investments. In the entire sample from 2004 to 2013, the sample contains 9,138 first-round investments which result from 4,685 unique start-ups and 1,167 unique investors that are based in the United States. Each start-up has a mean of 1.95 investors, while each investor has a mean of 7.83 investments. In order to have a sample where the five and seven calendar year exit windows are applied evenly for all observations, I drop all investments that occur after 2008 (2006) for the sample used to examine the five (seven) calendar year exit windows, since an investment that occurred in 2009 would have less than five calendar years to exit by the end of the sample in 2013. The mean number of investors for each start-up is relatively similar for all three samples, but the mean number of investments by each investor is smaller for the exit samples. In particular, the standard deviation for the mean number of investments by an investor increases as the sample period is extended, suggesting that investors are more likely to invest in a wider

variety of start-ups over time. Table 1.1 Panel B reports the number of investments that exited through either an acquisition or IPO for the exit samples. 21.74% (32.26%) of the investments in the five (seven) calendar year exit sample exited within the five (seven) calendar year window. The percentage of exits for the seven calendar year exit sample is larger, probably because of the longer time window.

In order to control for geographical effects, I classify the start-ups and investors into six geographical areas based on their zip codes, namely: Boston, Chicago, New York, Seattle, San Francisco / Silicon Valley, and the Rest of the United States. The first four geographical areas were chosen on the basis of their reputation as areas of high start-up activity and where start-ups are more likely to be located at. Finally, since Crunchbase does not provide industry classifications for start-ups, I implement the supervised learning method of Gradient Boosted Trees (Friedman, 2001; 2002) to classify the start-ups into industry groups. Specifically, I use the start-up industry categories and descriptions from Thomson Reuters VentureXpert to train an algorithm to classify the start-ups in Crunchbase to industry categories. The resulting algorithm was able to predict the training sample from VentureXpert correctly about 97% of the time, but was only able to predict a validation sample from VentureXpert correctly about 70% of the time.

Finally, since the Crunchbase data contains only actual matches between start-ups and investors, statistical testing of the odds of being matched is not possible with the available data as is. In order to overcome this, I propose several counter-factual investors for each actual investment. I use two methods for this: first, by finding investors that are most similar to the investor that did the actual investment in terms of data on their investment preferences where available in VentureXpert; and second, by randomly assigning other investors in the Crunchbase data as a counter-factual investors. I provide more details on the creation of counter-factuals in the methodology section.

1.2.2 Educational similarity

Using the data on start-up and start-up investor individuals in Crunchbase, I identify the individuals associated with each investment. I filter the start-up individuals for founders and entrepreneurs, while filtering the investor individuals for founders, principals, and partners. I term these individuals as start-up entrepreneurs and investor principals. Where possible, I include only individuals who are present at the associated start-up or investor for at least six months before the investment, so as to ensure that only individuals who were involved with the investment are included in the sample.

Using the identified individuals for each investment, I create all possible combinations of the start-up entrepreneurs and the investor principals. I examine the tertiary educational institutions attended by each combination of a start-up entrepreneur and an investor principal. If at least one start-up entrepreneur and investor principal pair for that investment attended the same institution, I designate that investment as having an educational institution similarity. As the educational institution names are not standardised in the available version of the Crunchbase data, the names were matched to a data-set of institution names based on a web crawl of LinkedIn and Webometric pages via a combination of both automated and manual matching. Specifically, the automated matching was first implemented, and then supplemented with manual matching until at least 90% of the reported institution names in Crunchbase were matched to a specific institution. I include foreign universities in my measure of educational similarities, since the aim is to capture educational similarity rather than simply American educational similarity.

Table 1.2 reports the twenty most common educational institution similarities among all the possible combinations of start-up entrepreneurs and investor principals that were involved in the same investment. For example, in the case of a start-up entrepreneur who attended Columbia, Yale, and Princeton and an investor principal who attended Columbia and Princeton, their connection would be designated as a educational institution similarity for both Columbia and Princeton. Stanford and Harvard make up 66.9% of the educational

institution similarities in the sample, with the Massachusetts Institute of Technology being the next most common with 6.3%. Public schools make up six spots of the top twenty. Additionally, the only non-American university in the top twenty is Tel Aviv University from Israel, which could also indicate similarities on other dimensions such as culture, heritage, and religion with regards to the ethno-religious group of Jewish people.

Following Cohen, Frazzini, and Malloy (2008), I also use the following degree types in the analysis: Bachelors, MBA, Masters, Law School, Medical School, or PhD. For example, in this case a degree similarity is present if the entrepreneur-principal pair both graduated with an MBA, regardless of educational institution. Table 1.3 presents the number of degree similarities between start-up entrepreneurs and investor principals affiliated with an investment. As expected, start-up entrepreneurs and investor principals involved in an investment are most likely to have both graduated with a Bachelors degree, followed by MBA degrees. In contrast, degree similarities based on Law and Medicine degrees make up only about 0.51% of all entrepreneur-principal degree similarities.

The measure of educational similarity is then enhanced by requiring both the degree types and educational institutions to match. For example, an educational similarity based on both educational institution and degree type occurs only if the entrepreneur-principal pair both graduated from Stanford with an MBA. The inclusion of the degree type increases the odds that an actual connection existing between the two individuals. Table 1.4 reports the twenty most common educational similarities based on both educational institution and degree type between start-up entrepreneurs and investor principals. Bachelor degrees are involved in half of the top twenty institution and degree connections, while Stanford is the most well represented institution, with Stanford Bachelors, Masters, MBAs, and PhDs all appearing on the top twenty list. This is possibly due to the proximity of Stanford to the San Francisco Bay Area and Silicon Valley. Only two American public school degrees, Berkeley and Michigan (both Bachelor degrees), appear in the top twenty. Tel Aviv University remains the only non-American institution to appear on the list.

Unfortunately, since graduation year data is incomplete in Crunchbase, I am unable to

implement the additional restriction on graduation year used by Cohen, Frazzini, and Malloy (2008), which would further increase the probability that the individuals are socially acquainted with each other. However, given that start-up entrepreneurs are generally relatively younger than investor principals, it is unlikely that there would be a significant number of matches with investor principals in terms of graduation year.

Using the various measures of educational similarity, I compute the measures of educational connection between the start-up entrepreneurs and investor principals involved in an investment. The educational institution connection (educational institution and degree type) measure takes the value of 1 if at least one pair of start-up entrepreneur and investor principal involved in an investment both graduated from the same educational institution (same educational institution and degree type), and 0 if no such pair exists. However, the degree type connection measure is computed slightly different, as the measure takes the value of 1 if at least one pair of start-up entrepreneur and investor principal both graduated with the same type of degree, excluding the Bachelor degrees, and 0 otherwise. I exclude Bachelor degrees since Bachelor degrees are relatively common within the sample. Table 1.5 shows that the presence of an educational institution connection only occurs in 31.9% of the investments, while the educational institution and degree connection only occurs in 20.3% of the investments. In contrast, a non-Bachelors degree connection occurs in 40.6% of the investments. Additionally, the proportion of investments with educational connections in the actual sample exceed those in both of the counter-factual samples. The difference is statistically significant, suggesting that the probability of investment is positively related with the presence of an educational similarity between the start-up entrepreneurs and investor principals.

1.2.3 Ethnicity and cultural distance

Existing methods of classifying individuals to ethnic groups are largely based on the family names. Bengtsson and Hsu (2015) base their ethnicity to name matching on the lists provided by Kerr (2008), which reports the 100 most common family names across the following

ethnic groups in the Melissa Data Corporation ethnic-name database: Chinese, English, European, Hispanic/Filipino, Indian/Hindi, Japanese, Korean, Russian, and Vietnamese. However, the Kerr lists appear to be inaccurate. For example, consider the following ethnicity classifications: Adler is Indian/Hindi, Laursen is Vietnamese, Eyüboğlu is Korean, Lee is Russian, Sigmund is Japanese, and Kulprathipanja is Hispanic/Filipino. These obvious mistakes, coupled with the relatively small number of family names considered, suggest that depending on the Kerr (2008) methodology is sub-optimal. In contrast, Hegde and Tumlinson (2014) rely on a vendor of proprietary name-based classification services, Origins Info, for ethnicity name-matching, but do not go into detail on how the matching is implemented. Pan, Siegel, and Wang (2016a, 2016b) use passenger lists of ships arriving in New York City from foreign ports between 1820 and 1957 that are obtained from Ancestry.com to build a frequency distribution of countries of origin for non-U.S. passengers with non-missing data.

I opt to use data from the Oxford Dictionary of American Family Names (ODAFN Hanks, 2003). Mateos (2007) examines the Cultural Ethnic Language Group (CELG) method outlined in Tucker (2005) that is implemented in the ODAFN, and suggests that the CELG method is superior to all other existing methods due to the design and its estimated accuracy of 88%-94%. However, the ODAFN does not identify a unique ethnicity for each surname, but instead identifies all possible ethnicities associated with a family name. This makes for a coarser measure of ethnicity, given that using name data to identify ethnicity is already imprecise. However, as long as the imprecision is randomly distributed, the estimates should not be biased.

I identify the ethnicities of each start-up entrepreneur and investor principal by examining the family names in conjunction with the ODAFN. I also consolidate several of the ethnicities listed in the ODAFN into broader categories, such as the English, Hispanic, Dutch, and Arab ethnicities. If at least one start-up entrepreneur and investor principal pair involved in an investment share at least one ethnicity based on the ethnicity profile of their last name, I code this as an ethnic match. Additionally, I create an alternative ethnicity measure that is similar to the prior measure, but limited to ethnicities identified by Bengtsson and Hsu

(2015) as being active in the start-up and venture capital industry in the United States, namely: Chinese; Indian; Japanese; Korean; Russian; Hispanic; Vietnamese; and Israeli / Jewish. While the Jewish people are not of an ethnic group, but rather tied by religion, culture, or heritage, Bengtsson and Hsu note that a large number of start-up entrepreneurs and investor principals belong to this group, suggesting a potential affinity effect among Jewish people in the sample.

Table 1.6 presents the number of ethnicity matches between entrepreneur-principal pairs involved in the same investment for each ethnicity. The table also reports the ethnicity similarity variables that I use for the analysis for the actual sample, and includes the numbers for the samples that include matched counter-factuals and randomised counter-factuals. While 80.4% of investments exhibit co-ethnicity, only 26.0% of all investments exhibit co-ethnicty based on the ethnicity sub-groups identified by Bengtsson and Hsu. The counter-factual samples both have a smaller proportion of investments that are connected by ethnicity. Additionally, the ethnic matches for all three samples are dominated by British, German, Irish, Israeli / Jewish, and French ethnic matches. With the exception of the Israeli / Jewish ethnic match, none of the other specific venture capital active ethnic matches appear in more than 4.5% of all investments. The proportion of ethnicity matches and active ethnicity subgroup matches are larger in the actual sample than both of the counter-factual samples, which again suggests that the probability of investment is positively related with the presence of co-ethnicity.

Next, I consider the cultural norms of start-up entrepreneurs and investor principals. I begin by using the possible ethnicities of each individual, which were identified through the ODAFN. The scores of the six Hofstede cultural dimensions were then calculated for each individual based on an equal weighting of all possible ethnicities for that individual. The mean cultural dimension scores for each start-up (investor) across all start-up entrepreneurs (investor principals) is then calculated, also on an equally weighted basis. While Karolyi suggests supplementing the Hofstede cultural dimensions with other variables from the World Values Survey (WVS), I find that the coverage of the WVS would lead to a significant

reduction in the number of available observations due to the difference in ethnicity coverage across the ODAFN, Hofstede, and the WVS. As such, I elect to use only the Hofstede measures, and create a measure of cultural distance between the start-up entrepreneurs and investor principals involved in an investment based on the following Euclidean distance calculation:

$$CD_{i,s} = \sqrt{\frac{1}{n}\sum_{d=1}^{n}(\frac{(I_{d,i} - I_{d,s})^2}{\sigma_d^2})} \tag{1.1}$$

where $I_{d,i}$ stands for the index for the d-th cultural dimension and i-th investor, $I_{d,s}$ stands for the index for the d-th dimension and s-th start-up, σ_d^2 is the variance of the index of the d-th dimension, n for the number of cultural dimensions, and $CD_{i,s}$ is the cultural difference of the i-th investor and the s-th start-up. A cultural distance of 0 indicates that the start-up entrepreneurs and investor principals are completely similar. In contrast, a cultural distance of 1 indicates that on average the two groups have cultural dimension scores that differ by one standard deviation of the scores.

Table 1.7 presents the summary statistics of the cultural distance measures in the sample of investments. The distribution of the cultural distance appears to be relatively consistent across the three samples, but the differences are statistically significant with the counter-factual samples showing a larger cultural distance compared to the actual sample, suggesting that the odds of an investment being made in a start-up is negatively related to the cultural distance between the start-up entrepreneurs and investor principals.

1.2.4 Gender

In order to identify the gender of the individuals in the sample, I use the Baby Names from Social Security Card Applications data provided by the Social Security Administration (SSA). The SSA data includes all baby names from Social Security card applications for births that occurred in the United States after 1879. While each name is identified for a certain gender, the proportion of males and females who share a given name is also provided.

I follow the method outlined by Blevins and Mullen (2015) in identifying gender based on the SSA data, and limit only use the data from 1940 to 2000 as this would be more likely to match the likely ages of the individuals in the database being classified.

For use in my analysis, I create a binary indicator for when an investment involves at least one female entrepreneur and at least one female investor principal. Table 1.8 shows that only 2.31% of investments in the actual sample fulfill this requirement. The proportion of investments where the requirements of this indicator is fulfilled in the counter-factual samples is similar to the original sample, but the differences between the actual and counter-factual samples are statistically insignificant. These numbers re-emphasise how uncommon it is to have a female presence on both sides of the investment.

1.3 Shared characteristics and selection

Do shared characteristics affect the mutual selection and matching between start-ups and start-up investors? While I can observe successful matches, it is not possible to observe the counter-factual of matches that did not come to fruition, such as an investor whose offer of funding was rejected by a start-up, or a start-up whose attempt to garner funding was rejected by an investor. This makes it difficult to analyse how the shared characteristics relate to the matching between start-ups and investors. My solution is to create counter-factual observations by selecting alternative investors for each start-up, either through a random draw or by matching on investor characteristics. I hypothesise that the odds of being matched increases in the level of similarities between the start-up and their investors.

1.3.1 Identifying the counter-factual

Every observed investment is likely to be accompanied by other unobserved investment offers that were rejected by the start-up. In order to test if shared characteristics between the start-up and investors affect the odds of being matched, I use two methods for creating possible counter-factual observations.

The first method draws on the similar approaches of Bena and Li (2014), Bengtsson and

Hsu (2015), and Hegde and Tumlinson (2014), where pseudo-counter-factual observations are created on the basis of certain criteria. For each investment, I find other investors from that are most similar to the actual investor for that start-up based on attributes other than the variables of interest in the analysis based on a brute nearest-neighbor algorithm. These investors, which I term pseudo-investors, take on the role of alternative investors for that start-up. The pseudo-investor and start-up pairs therefore serve as counter-factual observations. However, an alternative source of investor attributes is needed, such as Thomson Reuters VentureXpert is required since Crunchbase has limited data on investor attributes. The available investor attributes in VentureXpert include the location, type of investor, and most importantly, their investment preferences over the start-up industry, start-up location, and investment stage. I find pseudo-investors that are as similar to the original investor as possible based on these attributes, and select five pseudo-investors for each start-up investment from the set of investors present in the VentureXpert data.

I choose five pseudo-investors for each investor in a start-up based on the VentureXpert data, making sure that the pseudo-investors do not repeat across all the investments in that start-up. For example, if a pseudo-investor is the most similar investor to the two investors for a start-up, I only include the pseudo-investor once in the set of ten pseudo-investors that are assigned to that start-up. I fill the resulting empty spot with the next most similar pseudo-investor in the combined set of pseudo-investors of the two investors. However, the final matched counter-factual sample may have less than five pseudo-investors for each investor. As the pseudo-investors are chosen from the set of investors in VentureXpert, some of the resulting start-up and pseudo-investors pairs may not be usable due to the lack of corresponding Crunchbase data for the pseudo-investor. Therefore, while five pseudo-investors were chosen from VentureXpert, there may be less than five pseudo-investors in the matched counter-factual sample. The similarity variables of interest are then re-calculated for all (i) start-up and investor and (ii) start-up and pseudo investor pairs in the matched counter-factual sample. The resulting matched counter-factual sample includes 9,138 actual investments and 22,088 counter-factual investments, which works out to about 2.41 counter-

factual observations for each actual investment, denoting that about half of the pseudo-investor pairs are eliminated due to a lack of Crunchbase data.

The second method I use is based on a random draw of pseudo-investors from the Crunchbase data. For each investment, I randomly draw five other unique investors from the set of all investors with available data in Crunchbase as pseudo-investors. I then calculate the similarity variables of interest for the start-up and investor pairs, as well as the start-up and pseudo-investor pairs. The final random matched sample includes 9,138 actual investments and 42,070 counter-factuals, such that each investment has 4.60 counter-factuals. The number of counter-factual investments is less than five due to the lack of available corresponding data in Crunchbase. For example, random selection could result in a pseudo-investor that has very little corresponding usable data in Crunchbase.

I then estimate the following logistic regression with a dependent binary variable that takes on the value of 1 if the start-up and investor pair is an actual investment pair, and 0 if the observation involves a pseudo-investor. I include fixed effects for start-up funding year, start-up industry, as well as geographic fixed effects based on the location for both the start-up and the investor or pseudo-investor. I also include a variable that indicates if both the start-up and investor are both located in the same geographical region from the following: Boston, Chicago, New York, Seattle, San Francisco / Silicon Valley, and the Rest of the U.S.A. Additionally, the standard errors are double clustered on start-ups and investors.

$$\begin{aligned} Matched_{i,s} = \alpha + \beta Similarity_{i,s} + YearFE_s + GeographyFE_i + GeographyFE_s \\ + IndustryFE_s + RegionMatchFE_{i,s} + \epsilon_{i,s} \end{aligned} \tag{1.2}$$

where $Matched_{i,s}$ refers to the binary variable defined previously. α refers to the baseline probability of investment, while $Similarity_{i,s}$ refers to the similarity variables between investor i and start-up s. $\epsilon_{i,s}$ denotes the error term.

1.3.2 Matched counter-factual sample

Specification Matched in Table 1.9 presents the results of the logistic regression of the similarity variables on the dependent variable that indicates that the observation is an actual investment. The regression includes start-up funding year, start-up industry, and geography fixed effects, and uses the matched counter-factual sample. With the exception of the *FemaleBoth* variable, all of the other similarity variables are statistically significant at either the 1% or 5% level, a result that suggests that similarities are positively related to the odds of the start-up and investor being matched.

For the educational similarity variables, a start-up whose entrepreneurs share the same educational institution affiliation as the investor principals investing in the start-up has 38.1% higher odds of being matched with an investor compared to a start-up whose entrepreneurs share no common educational institution affiliation as the start-up investors. As the educational similarity measure based on the same institution and degree type can only take on the value of 1 if the institution only educational similarity variable also has the value of 1, I interpret the results as follows: the start-up whose entrepreneurs share an institutional affiliation and type of degree with its investor principals has 42.0% higher odds of obtaining funding from an investor compared to a start-up whose entrepreneurs only have an educational institution connection with its investor principals, suggesting that a stronger connection is positively related with the odds of being funded by the investor. Similarly, the odds of being funded or matched increases by 38.6% when at least one pair of start-up entrepreneurs and investor principals involved in the investment both have the same advanced degrees.

I also find that the co-ethnicity variables are positively related to the odds of the start-up and investor being matched. A start-up whose entrepreneurs share an ethnic similarity to at least one of their investor principals has 44.3% higher odds of being funded by the investor than a start-up whose entrepreneurs do not have a similar ethnicity as their investor principals. The odds of being matched with an investor increases when the co-ethnicity

between the start-up entrepreneurs and investor principals involves ethnic groups that were identified as being active in venture capital and start-ups by Bengtsson and Hsu (2015). Specifically, start-ups whose personnel share an ethnic sub-group similarity with at least one of the investor principals have 51.2% higher odds of being matched with an investor compared to a start-up whose personnel only have a general ethnic similarity with at least one of their investor principals. These results suggest that similarities in the educational and ethnic social network are positively related to the odds of a start-up being matched with an investor, possibly due to an expanded set of known investment opportunities.

The measure of cultural distance between the start-up entrepreneurs and the investor principals also suggest that similarities increase the odds of being successfully matched with an investor. Specifically, cultural distance is negatively related to the odds of a match between the start-up and the investor. The odds of the start-up being matched with the investor falls by 14.4% $(1 - 0.856)$ when the cultural distance between the start-up entrepreneurs and the investor principals increases by one standard deviation. This suggests that controlling for ethnicity, differences in cultural values between the entrepreneurs and investor principals may be considered detrimental to how the partnership might operate after being formed.

I find little evidence that the presence of female entrepreneurs and female investor principals in the investment is strongly related to the odds of matching between the start-up and the investor. Despite the presence of start-up and investor geography controls, I find that the *RegionMatch* variable that indicates if the start-up and investor are both located in the same American region is positively related to the odds of matching and is statistically significant. This suggests a co-region effect that may be due to the reduced travel and monitoring costs. Overall, these results suggest that similarities between start-up entrepreneurs and investor principals are positively related to the odds of the start-up being matched with an investor successfully.

1.3.3 Randomised counter-factual sample

The method used for creating the matched counter-factual sample may result in a biased sample due to the characteristics used to select the pseudo-investors. As an alternative, I instead randomly select pseudo-investors for each investment. Specification Randomised in Table 1.9 describes the results of the logistic regression of the similarity variables on the dependent variable that indicates if the observation is an actual investment with start-up funding year, start-up industry, and geography fixed effects, using the randomised counter-factual sample.

The results for the randomised counter-factual sample are similar to those of the matched counter-factual sample. Specifically, the magnitude of the coefficients are relatively similar for all variables, which supports the prior interpretation that similarities between the start-up entrepreneur and the investor principals are positively related to the odds of the start-up being matched with an investor. Additionally, the odds ratios in this regression are more statistically significant and the pseudo R-square is larger, which could be due to the difference in sample size between the two counter-factual samples.

However, the *RegionMatch* variable is significantly different in terms of magnitude with this sample compared to the matched counter-factual sample. This could be due to the how the matched counter-factual sample was built. As the pseudo-investors in the matched counter-factual sample were selected on the basis of being from the same geographical location in the U.S. as the original investor, there is already some form of control for region in the regression that will not be present in the randomised sample whose pseudo-investors were randomly chosen. This leads to a greater statistical weight on the *RegionMatch* variable in the regression of the randomised sample.

Overall, the results from the randomised counter-factual sample further support the hypothesis that similarities between start-up entrepreneurs and investor principals are positively related with the odds of matching between start-ups and investors, and that the odds of matching increases in the depth of the similarity. Additionally, I find results that are

consistent with those of Bengtsson and Hsu (2015) and Hegde and Tumlinson (2014), who find that co-ethnicity is positively related to the odds of the start-up and investor being matched together.

1.4 Shared characteristics and start-up performance

Given that shared characteristics increase the odds of matching, do these characteristics affect the performance outcomes of the start-ups, conditional on the start-ups having been matched to an investor? It is unclear which direction the effect of shared characteristics might have on start-up success. Shared characteristics can improve information sharing, communication, or decision making, such that start-ups are more likely to succeed. On the other hand, start-ups and investors may decide to work with others who are similar to themselves in spite of negative performance implications. For example, an investor may choose to fund start-up entrepreneurs belonging to the same ethnicity if it enhances the investor's standing within the ethnic community, even if most of the start-ups are inferior. In these cases, the investor may also be less demanding of similar start-ups after having been matched or be more lenient in terms of monitoring their investment such that the odds of the start-up succeeding is reduced.

1.4.1 Acquisitions and IPOs of start-ups

I estimate the same regression structure used in the analysis of matching between start-ups and investors, but with the following changes: first, I limit the sample to actual investments only; and second, I use a measure of start-up success as the dependent variable. For the dependent variable, I use two binary measures that take on the value of 1 when the start-up exits either by acquisition or IPO within five or seven calendar years of the first round of funding, and 0 otherwise. The choice of the measure of success is because data on the rate of return of the start-up investments is not available, and information on the funding raised by start-ups at all rounds is also similarly limited. As such I choose to use acquisitions or IPOs, as these events are clear signals that a start-up has succeeded. I limit the sample to

investments that have had the full five or seven calendar year window to exit in order to evaluate all actual investments on a consistent basis. Specifically, I estimate the following logistic model, with standard errors that are double clustered on start-ups and investors:

$$\begin{aligned} Exit_{i,s} = \alpha + \beta Similarity_{i,s} + YearFE_s + GeographyFE_i \\ + GeographyFE_s + IndustryFE_s + RegionMatchFE_{i,s} + \epsilon_{i,s} \end{aligned} \tag{1.3}$$

where $Exit_{i,s}$ refers to the binary variable of exit defined previously, either with a five or seven calendar year window. α refers to the baseline probability of investment, while $Similarity_{i,s}$ refers to the similarity variables between investor i and start-up s. $\epsilon_{i,s}$ denotes the error term.

Table 1.10 presents the results of the logistic regression of the similarity variables on start-up success. Start-up success in this instance is defined as whether the start-up exits within five or seven calendar years of the first round of funding by the investor, conditional on the start-up having already being matched with and obtaining funding from the investor. The regression includes start-up industry, funding year, and location fixed effects. In contrast to the previous regressions that examine the odds of being matched, the odds ratios of the similarity variables from the regressions are largely consistent with the interpretation that similarities between the start-up entrepreneurs and the investor principals are negatively related to the odds of the start-up exiting within either five or seven calendar years after the first round of funding.

While the results are largely statistically insignificant, there are some variables that are statistically significant at the 5% level only for the regression on the exit variable that is based on a five calendar year window, but not for the seven calendar year window. The presence of a higher level degree similarity between the start-up entrepreneurs and the investor principals relates to a 26.4% $(1 - 0.736)$ reduction in the odds of the start-up exiting within five years, while an ethnicity match is related with a 28.3% $(1 - 0.717)$ reduction in the odds of the start-up exiting within five years. Similarly, an increase of one standard deviation in the cultural distance measure between the start-up entrepreneurs and the investor principals

relates to a 41.4% increase in the odds of the start-up exiting within five years. These results suggest that differences, not similarities, are positively related to the chances of the start-up exiting within five years of being funded for the first time. These results contradict those of Hedge and Tumlinson (2014), who find that co-ethnicity increases the probability that the start-up exits the start-up stage by either being acquired or going public.

Further, I find limited evidence that the presence of at least one female start-up entrepreneur and at least one investor principal in an investment is negatively related (47.8% $(1 - 0.522)$, 10% significance level) to the odds of the start-up exiting within seven years of the first round of funding. This is in contrast to the results presented by Raina (2017), who reports that female-led start-ups are more likely to exit if the start-up has a female investor. This difference in results could be due to the variable that I use, which is based on all female start-up entrepreneurs and executives. In contrast, Raina only examines female leaders of start-ups. The effect I observed could therefore be diluted in comparison to that of Raina.

Additionally, the *RegionMatch* variable is negatively related (20% $(1 - 0.800)$ or 19.7% $(1 - 0.803)$) to the odds of the start-up exiting within either five or seven years after the first round of funding. This suggests that the start-up being located in the same region as the investor is a detriment to the performance of the start-up, potentially because the start-up has to deal with an increased level of scrutiny from the nearby investor that interferes with business operations, or that the investor is more likely to actively monitor their start-ups that are located further away once the funding has been given. Alternatively, it could also indicate that these cross-region investments are due to start-up investors that have access to a broader network, and can hence invest in other geographical regions.

Overall, the evidence suggests that, conditional on the start-up being selected and matched to an investor, similarities between the start-up personnel and investor principals are at worst negatively related and at best unrelated to the odds of the start-up exiting the start-up stage within five or seven years of the first round of funding,

1.4.2 Heckman selection correction

The prior estimations lend themselves to the hypothesis that similarity between start-up entrepreneurs and investor principals is negatively related to start-up performance, as measured by exits within a five or seven calendar year window, conditional on the start-up having been matched with an investor. Since the identified effect is based on conditional correlations, it is not clear if the effect is due to how the investor selects what start-up to invest in or due to the similarity between the start-up entrepreneurs and investor principals. An interpretation consistent with the former is that while high quality start-ups have no shortage of willing investors, lower quality start-ups have to work hard to obtain funding, and are more likely to tap into their connections for investors and funding. This could explain why I observe a negative relation between the similarities and the odds of start-up exit.

In order to correct for the potential bias that results from the selection process I follow the approach outlined in Heckman (1979) of estimating a first-stage selection equation, which is then used in the estimation of the second-stage outcome equation. However, there must be at least one variable that appears in the selection equation but not in the outcome equation in order to generate credible estimates. In the context of this setting of start-ups and investors, there must be at least one variable that is related to the odds of the start-up being matched to an investor, but unrelated to the odds of start-up exit.

The instruments I use to fulfill the exclusion requirement are based on the percentage of similar investments within the market of each start-up, which I define as the industry, year, and region that the start-up belongs to. Since the variables are proxies for the availability of similar partners within the market of the start-up, the variables will also be related to the odds of a start-up being matched with a similar investor. However, the availability of similar partners within the market of the start-up is unlikely to be related to the odds of start-up exit. While an ethnic Chinese investor may be more likely to encounter entrepreneurs of Chinese ethnicity in San Francisco compared to Boston, the Chinese investor in San Francisco is unlikely to have an advantage over the Chinese investor in Boston in the evaluation of

co-ethnic entrepreneurs through ethnicity based channels. Additionally, it is unlikely that the average quality of similar investments made by the investor (and hence the start-up performance) will differ across regions, unless (i) the investor has differing preferences for similar investments across the very same regions, and / or (ii) the investor uses different criteria or thresholds to initiate investment in different regions.

The instruments are generally computed as follows. Starting with the sample of actual and matched counter-factual investments, I identify the unique industry-year-region markets in the data, such as Healthcare-2005-Seattle. For each market, I find all of the investments and counter-factual investments, and then identify those where the start-up entrepreneurs and investor principals share characteristics with each other. The variable is then the ratio of the number of similar investments or pseudo-investments in that market to the number of all investments and pseudo-investments in that market. For example, if the Healthcare-1999-Seattle market has 6 similar investments / pseudo-investments out of 12 investments / pseudo-investments, the percentage of similar investments in that market is 50%. I repeat this calculation for all of the similarity variables of education, ethnicity, cultural distance, and gender, such that I have an instrument for each of the similarity variables.

Table 1.11 reports the OLS estimates of the Heckman selection correction for the five and seven calendar year exit windows. With the exception of the female variable, the results for the five calendar year exit window support the hypothesis that similarity has a negative causal effect on the likelihood of a start-up exit after controlling for potential selection bias. When the exit window is lengthened to seven calendar years, the results based on the sample with randomised counter-factuals is similar but with diminished magnitudes and significance, while most of the variables appear to no longer be related to the likelihood of start-up exits in the case of the sample with matched counter-factuals. However, in both samples with seven calendar year exit windows, the presence of a female connection on both sides of the investment has a negative and significant effect on start-up exits, which may denote that the start-up sector is still largely male-dominated, and that this dominance has spillover effects on performance.

Overall, the estimates lend themselves to the conclusion that there is a negative effect of similarity on start-up performance, but also that the effect diminishes in magnitude over time. Additionally, while the Heckman estimates are not easily cross-comparable to those of the simple logit models, the similarity variables are generally more likely to be statistically significant when the Heckman selection correction has been applied, suggesting that there is selection bias that was not controlled for in the logit models.

1.5 Conclusion

I examine how the shared characteristics between start-up entrepreneurs and their investor principals relate to the odds of two distinct start-up milestones: start-up funding and exit. I use Crunchbase, a novel database of start-ups, start-up investors, and associated personnel to build several variables that measure similarities in personal characteristics that are based on three categories: network (education and ethnicity); social values (cultural values); and both (gender). I find evidence that educational similarity is positively related to the odds that a start-up and investor partner together by up to 96.1% $((1.381 \times 1.420) - 1)$ depending on the depth of educational similarity in a matched sample and up to 143.0% $((1.651 \times 1.472) - 1)$ in a randomised sample.

However, educational similarity reduces the odds that the start-up exits within five (seven) calendar years by -23.6% to -34.1% (-21.3% to -22.9%). I also find that while ethnic similarities between the start-up entrepreneur and investor principal are positively related to the odds of matching (44.3% and 38.5%), and that the positive relation is strengthened if the ethnic similarities are based on ethnicities that have been identified as being particularly active in the venture capital industry (118.2% $((1.443 \times 1.512) - 1)$ and 235.1% $((1.385 \times 1.698) - 1)$). I also implement the Heckman selection correction in order to correct for selection bias and to better isolate the effect of similarity on the likelihood that a start-up exits. The Heckman models show that the similarity variables are generally negatively related to the likelihood of start-up exits, lending further support to the evidence that similarity is not optimal for start-up performance.

These results suggest that while start-ups and start-up investors are more likely to form partnerships with others who are similar to themselves, but that these partnerships are also less likely to succeed. Additionally, the results indicate that similarities should perhaps be avoided. This is in contrast to the findings of Hegde and Tumlinson (2014), who find that co-ethnicity is positively related to the likelihood of matching between start-up and investor, as well as the likelihood that the start-up exits. In contrast, my results appear to be somewhat contradictory. Given that similar start-ups are less likely to perform well, why would investors be more likely to invest in these start-ups? Perhaps investor principals could have overvalued the potential synergy gains with similar entrepreneurs, or are over-confident in their ability in assessing the true quality of similar start-ups. Another possible explanation is that investors enjoy working with entrepreneurs who share similarities with them, and are willing to forgo performance to do so, possibly by reducing their thresholds for investment.

Table 1.1: Start-ups, investors, investments, and investment outcomes

This table presents the number of start-ups, investors, investments, and investment outcomes five or seven calendar years after the first funding round for each investment in the sample of all investments between 2004 to 2013. The investments consist of investments made in start-ups based in the USA by investors that are also based in the USA. Panel A reports the number of investments, unique start-ups, and unique investors for three samples: the full sample from 2004 to 2013, and two other samples from 2004 to 2008 and 2004 to 2006. The two smaller samples are restricted by year in order to allow all of the investments to have the same time window to exit within five or seven years of the first funding round before the end of the sample. The mean and SD reported denote the number of participants from the other side of the partnership. For example, the mean reported for the start-ups denotes to the mean number of investors for each start-up, while the mean reported for the investors denotes the mean number of start-ups invested in by each investor. Panel B reports the number of investments that exited within five years (2004 to 2008) or seven years (2004 to 2006) of the first funding round.

Table 1.1: Start-ups, investors, investments, and investment outcomes

Panel A: Investments, startups, and investors

	2004-2013			2004-2008			2004-2006		
	N	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
Start-ups	4,685	1.950	1.248	2,018	2.027	1.274	1,023	2.212	1.319
Investors	1,167	7.830	14.445	788	5.190	7.408	584	3.875	4.922
Investments	9,138	-	-	4,090	-	-	2,263	-	-

Panel B: Investment outcomes

	2004-2008		2004-2006	
	N	**%**	**N**	**%**
Exits	889	21.74%	730	32.26%
Investments	4,090	100.00%	2,263	100.00%

Table 1.2: Most common institution-only educational connections of the investments

This table presents the twenty most common institution-only educational connections between the start-up executive and investor principals for all investments between 2004 to 2013. An institution-only educational connection exists if at least one start-up executive and one investor principal graduated from the same educational institution.

Table 1.2: Most common institution-only educational connections of the investments

Rank	Institution	N	%
1	Stanford University	8,722	41.20%
2	Harvard University	5,442	25.70%
3	Massachusetts Institute of Technology	1,334	6.30%
4	University of California Berkeley	1,120	5.29%
5	University of Pennsylvania	793	3.75%
6	Tel Aviv University	257	1.21%
7	Columbia University	245	1.16%
8	University of Michigan	232	1.10%
9	Dartmouth College	208	0.98%
10	University of California Los Angeles	196	0.93%
11	Cornell University	183	0.86%
12	Northwestern University	177	0.84%
13	Yale University	148	0.70%
14	University of Texas Austin	126	0.60%
15	University of Illinois Urbana Champaign	100	0.47%
16	New York University	97	0.46%
17	University of Chicago	93	0.44%
18	University of Virginia	92	0.43%
19	Duke University	87	0.41%
20	Princeton University	84	0.40%
	Total	21,172	100.00%

Table 1.3: Degree-only educational connections of the investments

This table presents the distribution of the degree-only educational connections between the start-up executives and investor principals for all investments between 2004 to 2013. A degree-only educational connection exists if at least one start-up executive and one investor principal graduated with the same type of degree from the following: Bachelors, Masters, MBA, Law, Medicine, and PhD.

Table 1.3: Degree-only educational connections of the investments

Degree Type	**N**	**%**
Bachelors	146,433	76.37%
Masters	12,175	6.35%
MBA	28,746	14.99%
Law	456	0.24%
Medicine	514	0.27%
PhD	3,428	1.79%
Total	191,752	100.00%

Table 1.4: Most common institution-and-degree educational connections of the investments

This table presents the twenty most common institution-and-degree educational connections between the start-up executives and investor principals for all investments between 2004 to 2013. A institution-and-degree educational connection exists if at least one start-up executive and one investor principal graduated from the same educational institution with the same type of degree from the following: Bachelors, Masters, MBA, Law, Medicine, and PhD.

Table 1.4: Most common institution-and-degree educational connections of the investments

Rank	Institution	Count	%
1	Harvard University - MBA	1,592	24.71%
2	Stanford University - MBA	925	14.36%
3	Stanford University - Bachelors	659	10.23%
4	Stanford University - Masters	415	6.44%
5	Harvard University - Bachelors	314	4.87%
6	University of California Berkeley - Bachelors	304	4.72%
7	University of Pennsylvania - MBA	148	2.30%
8	University of Pennsylvania - Bachelors	116	1.80%
9	Massachusetts Institute of Technology - Bachelors	105	1.63%
10	Massachusetts Institute of Technology - Masters	87	1.35%
11	Tel Aviv University - Bachelors	77	1.20%
12	University of Michigan - Bachelors	75	1.16%
13	Massachusetts Institute of Technology - PhD	73	1.13%
14	Cornell University - Bachelors	64	0.99%
15	Yale University - Bachelors	64	0.99%
16	Dartmouth College - Bachelors	61	0.95%
17	Columbia University New York - MBA	59	0.92%
18	Stanford University - PhD	59	0.92%
19	University of Chicago - MBA	58	0.90%
20	Princeton University - Bachelors	51	0.79%
	Total	6,443	100.00%

Table 1.5: Educational similarity of the investments

This table presents the number of investments from 2004 to 2013 that are connected through the educational similarity variables. *EducationInst* refers to the institution-only educational connection between the start-up executives and investor principals, while *EducationInstDeg* refers to the institution and degree educational connection between the start-up executives and investor principals. *EducationDegHigher* refers to the non-Bachelor-degree-only educational connection between the start-up executives and investor principals. The table also presents the difference in the proportion of the educational matches between the actual sample and the matched or randomised samples. ***, **, and * indicate significance at the 1%, 5%, and 10% levels respectively.

Table 1.5: Educational similarity of the investments

	Actual		**Matched**		**Randomised**	
Variable	**N**	**%**	**N**	**%**	**N**	**%**
EducationInst	2,917	31.92%	8,326	22.95%	8,566	16.73%
EducationInstDeg	1,855	20.30%	4,759	13.12%	4,653	9.09%
EducationDegHigher	3,716	40.67%	12,765	35.19%	14,738	28.78%
Total	9,138	100.00%	36,274	100.00%	51,208	100.00%

	Matched		**Randomised**	
Variable	**Diff. Prop.**	**t-stat**	**Diff. Prop.**	**t-stat**
EducationInst	0.090***	16.753	0.152***	29.512
EducationInstDeg	0.072***	15.725	0.112***	25.511
EducationDegHigher	0.055***	9.575	0.119***	21.551

Table 1.6: Ethnicity matches of the investments

This table presents the number of investments from 2004 to 2013 with ethnicity matches between start-up executives and investor principals. *EthnicityMatch* denotes the presence of an ethnic match between the start-up executives and investor principals, while *EthnicitySGMatch* denotes the presence of an ethnic match between the start-up executives and investor principals for the following ethnicities: Chinese; Indian; Japanese; Korean; Russian; Hispanic; Vietnamese; and Israeli / Jewish. The table also presents the difference in the proportion of the ethnicity matches between the actual sample and the matched or randomised samples. ***, **, and * indicate significance at the 1%, 5%, and 10% levels respectively.

Table 1.6: Ethnicity matches of the investments

	Actual		Matched		Randomised	
	N	%	N	%	N	%
Arab	31	0.34%	53	0.15%	67	0.13%
Armenian	1	0.01%	1	0.00%	1	0.00%
British	5,591	61.18%	20,820	57.40%	27,903	54.49%
Chinese	372	4.07%	930	2.56%	1,101	2.15%
Croatian	1	0.01%	3	0.01%	3	0.01%
Czech	48	0.53%	107	0.29%	129	0.25%
Danish	88	0.96%	251	0.69%	297	0.58%
Dutch	585	6.40%	1,780	4.91%	2,136	4.17%
Filipino	1	0.01%	1	0.00%	3	0.01%
French	1,025	11.22%	3,288	9.06%	3,761	7.34%
German	3,441	37.66%	12,907	35.58%	15,650	30.56%
Greek	32	0.35%	72	0.20%	81	0.16%
Hispanic	39	0.43%	90	0.25%	92	0.18%
Hungarian	20	0.22%	61	0.17%	61	0.12%
Indian	411	4.50%	926	2.55%	944	1.84%
Irish	1,838	20.11%	6,046	16.67%	7,079	13.82%
Israeli / Jewish	1,729	18.92%	5,030	13.87%	6,153	12.02%
Italian	345	3.78%	1,000	2.76%	1,114	2.18%
Japanese	4	0.04%	18	0.05%	13	0.03%
Korean	87	0.95%	179	0.49%	205	0.40%
Norwegian	62	0.68%	142	0.39%	177	0.35%
Polish	87	0.95%	166	0.46%	203	0.40%
Portuguese	23	0.25%	77	0.21%	89	0.17%
Russian	0	0.00%	1	0.00%	2	0.00%
Serbian	1	0.01%	3	0.01%	1	0.00%
Slovak	18	0.20%	35	0.10%	46	0.09%
Slovenian	8	0.09%	30	0.08%	30	0.06%
Swedish	65	0.71%	181	0.50%	249	0.49%
Ukrainian	1	0.01%	1	0.00%	1	0.00%
Vietnamese	2	0.02%	6	0.02%	11	0.02%
EthnicitySGMatch	2,375	25.99%	6,637	18.30%	7,945	15.52%
EthnicityMatch	7,348	80.41%	27,488	75.78%	36,921	72.10%
Investments	9,138	100.00%	36,274	100.00%	51,208	100.00%

	Matched		Randomised	
Variable	**Diff. Prop.**	**t-stat**	**Diff. Prop.**	**t-stat**
EthnicitySGMatch	0.077***	15.334	0.105***	21.557
EthnicityMatch	0.046***	9.810	0.083***	18.065

Table 1.7: Cultural distance of the investments

This table presents the average cultural distance for the start-up executives and investor principals of all actual investments from 2004 to 2013. The cultural distance (*CulturalDistance*) is calculated as a Euclidean distance based on the six Hofstede cultural dimensions. The table also presents the difference in the average cultural distance between the actual sample and the matched or randomised samples. ***, **, and * indicate significance at the 1%, 5%, and 10% levels respectively.

Table 1.7: Cultural distance of the investments

	Actual	**Matched**	**Randomised**
CulturalDistance	0.887	0.940	0.983
		Diff..	**t-stat**
Actual-Matched		-0.053***	-10.065
Actual-Randomised		-0.095***	-18.803

Table 1.8: Female presence within the investments

This table presents the proportion of investments where there is both a female start-up entrepreneur and a female investor principal. *FemaleBoth* is positive if there is at least one female in the ranks of both the start-up executives and investor principals, and zero otherwise. The table also presents the difference in the proportion of investments with this female representation between the actual sample and the matched or randomised samples. ***, **, and * indicate significance at the 1%, 5%, and 10% levels respectively.

Table 1.8: Female presence within the investments

	Actual		**Matched**		**Randomised**	
Variable	**N**	**%**	**N**	**%**	**N**	**%**
FemaleBoth	211	2.31%	896	2.47%	1,167	2.28%
Total	9,138	100.00%	36,274	100.00%	51,208	100.00%
			Matched		**Randomised**	
Variable			**Diff. Prop.**	**t-stat**	**Diff. Prop.**	**t-stat**
FemaleBoth			-0.002	-0.910	0.000	0.177

Table 1.9: Probability of matching between start-up and investor

This table presents the logistic estimation of the similarity variables on the likelihood of the start-up investment being successfully matched with an investor. For each actual start-up investment, pseudo-investors were chosen using two methods: first, five pseudo-investors were selected based on similarities in investor preferences and characteristics from Thomson Reuters VentureXpert; and second; a random draw of five investors from the set of investors in the actual sample. *EducationInst* refers to the institution-only educational connection between the start-up executives and investor principals, while *EducationInstDeg* refers to the institution and degree educational connection between the start-up executives and investor principals. *EducationDegHigher* refers to the non-Bachelor-degree-only educational connection between the start-up executives and investor principals. *EthnicityMatch* denotes the presence of an ethnic match between the start-up executives and investor principals, while *EthnicitySGMatch* denotes the presence of an ethnic match between the start-up executives and investor principals for the following ethnicities: Chinese; Indian; Japanese; Korean; Russian; Hispanic; Vietnamese; and Israeli / Jewish. *FemaleBoth* is positive if there is at least one female in the ranks of both the start-up executives and investor principals. *CulturalDistance* refers to the cultural distance calculated as a Euclidean distance based on the six Hofstede cultural dimensions. *RegionMatch* takes the value of 1 if the start-up and investor both belong to the same region out of the following regions (and 0 otherwise): Boston; Chicago; New York; San Francisco / Silicon Valley; Seattle; and Rest of USA. Standard errors are clustered at the start-up and investor level. Coefficients listed are odds ratios, while Z-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels respectively.

Table 1.9: Probability of matching between start-up investment and investor

	Matched	**Randomised**
EducationInst	1.381***	1.651***
	(4.48)	(9.97)
EducationInstDeg	1.420***	1.472***
	(4.95)	(6.46)
EducationDegHigher	1.386***	1.217***
	(4.99)	(5.25)
EthnicityMatch	1.443***	1.385***
	(4.49)	(5.48)
EthnicitySGMatch	1.512***	1.698***
	(5.73)	(9.63)
CulturalDistance	0.856**	0.871**
	(-2.26)	(-2.56)
FemaleBoth	0.976	1.033
	(-0.12)	(0.26)
RegionMatch	1.231***	3.826***
	(3.73)	(21.42)
Geographic controls	YES	YES
Year controls	YES	YES
Start-up industry controls	YES	YES
N	31,226	51,208
pseudo R-square	0.040	0.125

Table 1.10: Probability of investment exit within five or seven calendar years of first funding round

This table presents the conditional logistic estimation of the similarity variables on the likelihood of the start-up investment exiting within five or seven calendar years of being matched with an investor. *EducationInst* refers to the institution-only educational connection between the start-up executives and investor principals, while *EducationInstDeg* refers to the institution and degree educational connection between the start-up executives and investor principals. *EducationDegHigher* refers to the non-Bachelor-degree-only educational connection between the start-up executives and investor principals. *EthnicityMatch* denotes the presence of an ethnic match between the start-up executives and investor principals, while *EthnicitySGMatch* denotes the presence of an ethnic match between the start-up executives and investor principals for the following ethnicities: Chinese; Indian; Japanese; Korean; Russian; Hispanic; Vietnamese; and Israeli / Jewish. *FemaleBoth* is positive if there is at least one female in the ranks of both the start-up executives and investor principals. *CulturalDistance* refers to the cultural distance calculated as a Euclidean distance based on the six Hofstede cultural dimensions. *RegionMatch* takes the value of 1 if the start-up and investor both belong to the same region out of the following regions (and 0 otherwise): Boston; Chicago; New York; San Francisco / Silicon Valley; Seattle; and Rest of USA. Standard errors are clustered at the start-up and investor level. Coefficients listed are odds ratios, while Z-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels respectively.

Table 1.10: Probability of investment exit within five or seven calendar years of first funding round

	Five Years	**Seven Years**
EducationInst	1.102	1.177
	(0.67)	(0.96)
EducationInstDeg	0.862	0.813
	(-0.94)	(-1.21)
EducationDegHigher	0.736**	0.869
	(-2.20)	(-0.85)
EthnicityMatch	0.717**	0.808
	(-2.17)	(-1.10)
EthnicitySGMatch	0.892	0.792
	(-0.87)	(-1.54)
CulturalDistance	1.414**	1.307
	(2.45)	(1.46)
FemaleBoth	0.660	0.522*
	(-1.58)	(-1.75)
RegionMatch	0.800*	0.803*
	(-2.43)	(-1.95)
Geographic controls	YES	YES
Year controls	YES	YES
Start-up industry controls	YES	YES
N	4,086	2,258
pseudo R-sq	0.050	0.050

Table 1.11: Heckman correction for selection bias of exits

This table presents the OLS estimates of the second stage of the Heckman selection correction of the relationship between the variable indicating if the start-up investment exits within five or seven years of being matched with an investor, after having corrected for the selection bias in the first stage. Both the samples of actual and matched counter-factual investments and actual and randomised counter-factual investments are used. The excluded variables used for identification in the Heckman correction estimates are proxies of the availability of similar partners for a start-up in the year-region-industry market of the start-up based on the various measures of similarity. *EducationInst* refers to the institution-only educational connection between the start-up executives and investor principals, while *EducationInstDeg* refers to the institution and degree educational connection between the start-up executives and investor principals. *EducationDegHigher* refers to the non-Bachelor-degree-only educational connection between the start-up executives and investor principals. *EthnicityMatch* denotes the presence of an ethnic match between the start-up executives and investor principals, while *EthnicitySGMatch* denotes the presence of an ethnic match between the start-up executives and investor principals for the following ethnicities: Chinese; Indian; Japanese; Korean; Russian; Hispanic; Vietnamese; and Israeli / Jewish. *FemaleBoth* is positive if there is at least one female in the ranks of both the start-up executives and investor principals. *CulturalDistance* refers to the cultural distance calculated as a Euclidean distance based on the six Hofstede cultural dimensions. *RegionMatch* takes the value of 1 if the start-up and investor both belong to the same region out of the following regions (and 0 otherwise): Boston; Chicago; New York; San Francisco / Silicon Valley; Seattle; and Rest of USA. *Lambda* refers to the Inverse Mills Ratio generated by the first stage of the estimation. Standard errors are clustered at the start-up and investor level. Z-statistics are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels respectively.

Table 1.11: Heckman correction for selection bias of exits

	Matched-5Y	**Random-5Y**	**Matched-7Y**	**Random-7Y**
EducationInst	-0.062*	-0.082**	0.010	-0.011
	(-1.74)	(-2.18)	(0.23)	(-0.24)
EducationInstDeg	-0.108***	-0.084***	-0.065	-0.083*
	(-2.74)	(-2.62)	(-1.32)	(-1.86)
EducationDegHigher	-0.128***	-0.100***	-0.055	-0.064*
	(-4.10)	(-4.52)	(-1.25)	(-1.95)
EthnicityMatch	-0.153***	-0.148***	-0.071	-0.093**
	(-3.91)	(-4.30)	(-1.46)	(-2.08)
EthnicitySGMatch	-0.117***	-0.101***	-0.078	-0.091**
	(-3.07)	(-3.37)	(-1.59)	(-2.39)
CulturalDistance	0.092***	0.093***	0.067**	0.069**
	(3.58)	(3.82)	(1.96)	(2.15)
FemaleBoth	-0.052	-0.081	-0.113*	-0.167***
	(-1.09)	(-1.90)	(-1.77)	(-2.89)
RegionMatch	-0.084**	-0.268***	-0.058*	-0.164*
	(-3.57)	(-3.84)	(-1.88)	(-1.91)
Lambda	-0.555***	-0.459***	-0.167	-0.245
	(-3.02)	(-3.47)	(-0.70)	(-1.48)
Geographic controls	YES	YES	YES	YES
Year controls	YES	YES	YES	YES
Start-up industry controls	YES	YES	YES	YES
N	4,090	4,090	2,263	2,263
pseudo R-sq	0.033	0.019	0.056	0.041

Chapter 2

INFORMATION DISSEMINATION THROUGH SOCIAL CONNECTIONS: EVIDENCE FROM HEDGE FUND HOLDINGS

Social connections between individuals, such as through a common educational background (Cohen, Frazzini, and Malloy, 2008; 2010), interactions at work (Hvide and Östberg, 2015), or among neighbours (Brown, Ivković, Smith, and Weisbenner, 2008), have been shown to affect investment decisions. In particular, Cohen, Frazzini, and Malloy (2008) show that mutual fund managers hold more stock in connected firms, and that these connected holdings out-perform unconnected holdings. Additionally, they find that returns are concentrated around corporate news announcements, which suggest that mutual fund managers have an informational advantage due to their social connections. However, Cohen et al. admit that their results do not conclusively show that mutual fund managers and firm managers share information through an active connection, and could instead be due to mutual fund managers having an insight on the business decision making process of the firm managers due to their common educational background. In this study, we aim to address this problem by (i) investigating whether social connections between hedge fund companies and their portfolio firms are an information channel for financial markets, and (ii) whether market critical information is transferred between connected individuals.

A recent insider trading case illustrates how valuable information sharing between hedge fund company managers and portfolio firm personnel can be. Raj Rajaratnam, co-founder of the Galleon Group hedge fund, was arrested on charges of insider trading in October 2009. Rajaratnam was found to have obtained inside information in April 2007 on Intel's upcoming earnings announcement from an Intel executive, Rajiv Goel, whom he had befriended while

they were studying at the University of Pennsylvania Wharton School together. Taking advantage of the information obtained from Goel, Rajaratnam switched the position of his funds in Intel from a short position to a long position prior to the earnings announcement. The switch resulted in the Galleon Group avoiding a $23.5 million loss and instead gaining a $36 million profit, or a turnaround of nearly $60 million.

We focus on social connections between individuals that are based on educational institutions, and examine the educational ties between key personnel of (i) hedge fund companies and (ii) firms in the hedge fund companies' portfolios that undertake mergers or acquisitions of other firms. We classify the equity and option positions held by each hedge fund company in each of its portfolio firms as a net call or net put position, and show that the direction of the positions held by hedge fund companies in connected portfolio firms can predict the direction of the merger or acquisition announcement returns of connected portfolio firms undertaking an acquisition.

This particular empirical setting is helpful for our analysis in several ways. First, the decision to initiate a merger or acquisition is the private decision of the portfolio firm, not the hedge fund company. Hedge fund companies are therefore unlikely to have an informational advantage in this setting from their regular business operations over other well connected market participants, such as other non-connected investment funds. Second, hedge fund companies are relatively unregulated and have access to one of the largest sets of investment opportunities among market participants. This allows hedge fund companies to invest in a more diverse range of asset classes that can be more speculative in nature, as well as be more flexible in taking advantage of time sensitive information and investment opportunities. Third, while a connection between the hedge fund manager and the portfolio firm executive may indicate that they both have a better understanding of how each other makes business decisions, this very same connection may not necessarily lead to information sharing. Examining a specific event, such as merger transactions, increases the likelihood that outcomes are related to event-specific information transfer between individuals.

Using hand-collected biographical data on hedge fund company managers and BoardEx

data on portfolio firms executives and directors, we identify hedge fund company managers and key portfolio firm personnel that share a common educational background. We classify portfolio firms that have such a connection with their hedge fund investors as connected firms, and those without as unconnected firms. Since these educational connections are typically created many years prior to the event that we are examining for evidence of information transfer, reverse causality concerns are mitigated as information transfer is unlikely to be the reason for the formation of these educational ties.

We first examine the net positions taken by the hedge fund companies in connected portfolio firms relative to the net positions in unconnected portfolio firms. We manually collect information on option holdings by the hedge fund companies from the original public 13-F filings from the SEC Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system. The options held by the hedge fund companies in each portfolio firm are then classified into net call or net positions based on the procedure laid out in Aragon and Martin (2012). We then use these positions held by the hedge fund companies to build aggregated portfolios of connected and unconnected firms. We find that a long-short strategy of taking long positions in the connected net call firms and short positions in connected net put portfolio firms generates up to 1.37% in abnormal returns per month or 17.74% per year. In contrast, the same long-short strategy based on unconnected net call and unconnected net put portfolio firms reports abnormal returns which are statistically insignificant, suggesting that hedge fund managers have an informational advantage with regards to connected portfolio firms.

Next, in order to establish active information transfer between portfolio firm executives and hedge fund managers, we examine portfolio firms that have undertaken a merger or acquired another firm. We use the merger or acquisition announcement returns to test for active information transfer, since the merger or acquisition announcement returns would be sensitive to the provision of time-critical and private information on the merger or acquisition. We find that a net put (call) position held by a hedge fund company in a connected portfolio firm is associated with a lower (higher) likelihood that the future merger or ac-

quisition announcement returns of the portfolio firm is positive (negative), and that the magnitude of the effect increases in the strength of the connection. Our results suggest that the direction of the net position taken by a hedge fund company in a connected portfolio firm can predict the future stock performance of the connected portfolio firm. Additionally, a merger or acquisition is a time-critical and private decision by the connected portfolio firm. It is therefore less likely that connected hedge fund company managers can predict the direction of the merger or acquisition announcement return simply because of their shared educational history with connected portfolio firm executives. In this setting it is more likely that hedge fund company managers make their investment choices based on event-specific information that they may have obtained through their educational connection with portfolio firm executives.

We show that the hedge fund companies in our sample behave similarly to other investors examined in the extant literature, in that the hedge fund companies invest more in connected firms and earn higher returns on their investments in connected firms. This suggests that that our results are unlikely to be uniquely due to the composition of our sample and are instead likely to be generalisable. We first contrast the portfolio allocation decisions made by hedge fund companies in connected and unconnected portfolio firms, and the performance of investment portfolios constructed based on the presence of a connection. We show that hedge fund company managers over-weight their portfolios towards connected firms, and that the over-weighting increases with the strength of the connection. Specifically, the over-weighting increases from 2.619 basis points with the weakest connection variable to 40.10 basis points with the strictest criteria for a connection. The over-weighting is even more pronounced when hedge fund company managers are connected with senior directors of portfolio firms rather than with junior directors. We then examine the performance of connected holdings versus unconnected holdings based on calendar time portfolio returns. We classify the stocks held by the hedge fund companies into connected and unconnected portfolios, and calculate the weighted monthly returns for the two aggregated portfolios. We find that a long-short strategy that involves taking long positions in connected portfolio

stocks and taking short positions in unconnected portfolio stocks generates 80 basis points in abnormal return per month. Additionally, the magnitude of the returns increases with the strength of the connection. We also contrast the performance of connected stocks held by the hedge fund companies versus connected stocks that are not held by the hedge fund companies. A long-short strategy of taking long positions in connected portfolio stocks and taking short positions in connected non-portfolio stocks would generate 0.72% in abnormal return per month, or 8.99% per year. Overall, the results suggest that hedge fund companies have an advantage in investing in connected firms.

This paper contributes to the literature in two ways. First, we provide additional evidence on how social networks and direct existing connections between investors and their investment choices can affect investment decisions and performance. In particular, we address the challenge posed by Cohen, Frazzini, and Malloy (2008) to distinguish if the effect observed in the extant literature is due to information transfer between agents, or the familiarity that agents have of the thought process of their counterpart through their shared educational background.

Second and more generally, we document informed trading by hedge fund companies. Our results substantiate the hypothesis that hedge fund companies have a comparative advantage in information gathering and can obtain market critical information before other market participants. Consistent with the findings of Aragon and Martin (2012) on informed option performance, we find that net call positions held by a hedge fund company in a connected portfolio firm are associated with positive future stock performance. We also examine the net positions of portfolio firms that implement a merger or acquisition, and find that the net call (put) positions held by the hedge fund company in a connected portfolio firm prior to the merger or acquisition are associated with a higher (lower) likelihood that the merger or acquisition announcement returns are positive. Our results support the hypothesis that hedge fund companies gather corporate information directly from their connections at their portfolio firms and use the information to inform their investment choices.

The rest of the paper is organised as follows. First, we provide an overview of the

data sources used in this paper, and outline how the connection variables are coded and created. Second, we examine potential information dissemination through the education connection. We use fund equity and options holding data to create an alternative proxy for the direction of the bets on portfolio firms taken by hedge fund companies, and examine how fund investments predict future stock performance both in general and for portfolio firms that are involved a merger or acquisition. Third, we establish that the investment behavior of the hedge fund companies in our sample are similar to those of other investors already examined in the extant literature, in order to mitigate concerns that our results are driven by the type of investor. Finally, we summarise our findings and conclude.

2.1 Data and summary statistics

We build our data from several different sources. For the hedge fund data, we begin with hedge fund characteristics and manager names from the Barclays Hedge Fund Database (BHFD). We choose this database because it has archival data on the biographical information of top officials and key fund managers since the first quarter of 2007, in contrast to other hedge fund databases that are only able to provide the most current data on key personnel. The archival format of the database also mitigates the potential survival bias of hedge funds and hedge fund companies in the data.

We first identify hedge fund companies among the fund holding companies in the Barclays Hedge Fund database. We follow the method adopted by Griffin and Xu (2009) and Massoud, Nandy, Saunders and Song (2011), and identify hedge fund companies based on the criteria that 50% of more of the fund holding company's assets must be managed through hedge funds. When fund asset data is unavailable from the BHFD, we rely on Internet sources such as Businessweek/Bloomberg Private Company Information and Form ADV from the U.S. Securities and Exchange Commission (SEC). If these sources are unable to provide additional information, we examine the self-reported business descriptions from the websites of the holding companies to determine if their main business lies in hedge fund management.

We then supplement our biographical data on the executives at the sample hedge fund

companies with hand-collected employment and educational history data of fund managers and analysts of all ranks from a variety of Internet sources, such as Bloomberg, LinkedIn, and company websites. In particular, we collect tertiary education data, including the degree-granting institution, degree type, graduation year, and employment data for the time periods during which the fund managers worked for each of the hedge fund companies in our sample. Unfortunately, due to time and resource constraints, we limit our manual data collection to stand-alone hedge fund companies with fewer than 500 individuals that are either currently or previously affiliated to the hedge fund company on LinkedIn. Additionally, we obtain educational information on the directors of the portfolio firm from the BoardEx database. Unfortunately, the availability of biographical data from the BHFD and the BoardEx databases limit our sample period from the first quarter of 2007 to the first quarter of 2015.

We rely on two different sources for hedge fund company option holdings. Our first source is the Thomson Reuters 13-F Institutional Holdings (S34) database, which provides information on the long equity holdings of hedge funds at the holding company level. Our second source is the set of raw Form 13-F filings hosted on SEC EDGAR, which not only includes long equity holdings but the call and put options holdings as well. We use the options to classify the net direction of the positions held by the hedge fund companies in their portfolio firms. Since only institutions that have total holdings exceeding $100 million are required to file the Form 13-F with the SEC, we are limited to 429 unique hedge fund companies in our sample.

Finally, we rely on stock returns from the Center for Research in Security Prices (CRSP), and portfolio firm characteristics from Compustat.

2.1.1 Education connection

Our key variable of interest is the educational connection between hedge fund company managers and key portfolio firm personnel, which we define as board members and senior corporate officers. We create all possible combinations of hedge fund company managers and key portfolio firm personnel pairs and define the educational connection variables based on

these pairs. First, our baseline indicator variable, *Connection1*, takes the value of 1 if at least one combination of key portfolio firm personnel and hedge fund managers graduated from the same degree-awarding institution, and 0 otherwise. This denotes the instance where the pair have graduated from the same educational institution, and proxies for the possibility that the pair are socially acquainted with each other.

Next, we reinforce the *Connection1* variable by implementing additional criteria. This serves as a means of distinguishing the strength of the connection. For *Connection2A*, we require that at least one pair of hedge fund manager and key portfolio firm personnel graduated from the same institution with the same type of degree, based on the following degree groups proposed by Cohen, Frazzini, and Malloy (2008): PhD, MBA, Law School, Medical School, general graduate degree, and general undergraduate degree. This increases the likelihood that the pair are socially acquainted with each other. Similarly, for *Connection2B*, we require that at least one pair of hedge fund manager and key portfolio firm personnel graduated from the same institution in the same year, so as to be reasonably certain that the pair were on campus at the same time and therefore might have had an opportunity to directly interact and form an actual social tie.

For *Connection3*, we require all of the preceding criteria to apply simultaneously. Specifically, at least one pair of key portfolio firm personnel and hedge fund managers must have (i) graduated from the same degree-awarding institution with (ii) the same type of degree in (iii) the same year. Finally, we define *Connection4* by building on the criteria of *Connection3*, but also include the requirement that the hedge fund company manager in the pair must also be a top executive at his or her hedge fund company. In order to minimise the possibility of reverse causality such that the educational connection was formed for the purpose of information sharing, we remove all connections where the date at which the connection is observed is less than two years after the graduation year of either individual of the hedge fund company manager and key portfolio firm personnel pair. For example, we remove connections observed in 2010 if the hedge fund company manager involved graduated in 2009.

2.1.2 Summary statistics

Table 2.1 Panel A presents the summary statistics for the portfolio holding characteristics over time. Our sample includes a total of 429 unique hedge fund companies over 2007 to the first quarter of 2015. Each hedge fund company holds on average 84 stocks with an average (median) value of equity holdings of $724 million ($205 million).

Panel B describes the characteristics of the portfolio holdings of the hedge fund companies in our sample. The number of unique stocks held by the hedge fund companies in our sample decreases from 4,167 in 2007 to 2,819 in the first quarter of 2015. However, the number of stocks which have at least one educational connection (*Connection1*) decreases from 2,613 in 2007 to 1,818 in the first quarter of 2015. In contrast, requiring the connection to overlap in time (*Connection2B*) reduces the numbers to 485 and 310 respectively, while the number of hedge fund company and portfolio firm pairs with the strictest and strongest connection (*Connection4*) decreases even further to 127 and 33 respectively.

Panel C summarises the characteristics of the personnel on both sides of the connection. Our sample contains 5,240 hedge fund company managers with a total of 8,095 degrees, as well as 25,718 portfolio firm managers and directors with a total of 41,530 degrees. The numbers suggest that the average person in our sample has more than one degree, and is therefore likely to have more than one educational affiliation through which he or she may be connected to others in the sample.

Panel D displays the breakdown by the earliest graduation year of the individuals in our sample. 48% of hedge fund company managers graduated in 2000 or later, while 72% of the portfolio firm directors and managers graduated before 1980. The difference in the distribution of graduation year for the hedge fund company managers and key portfolio firm personnel suggest that the key portfolio firm personnel in our sample are significantly older than their hedge fund counterparts. It is therefore unlikely (on average) that the hedge fund company managers are hired primarily on the basis their connections with portfolio firm directors and managers.

Finally, Panel E displays the five most common educational institutions in our sample. The University of Pennsylvania, Harvard University, and Columbia University are the only institutions that rank among the five most common educational institutions for both hedge fund managers and portfolio firm directors. However, these institutions do not dominate the sample since only 22% of hedge fund managers and 17% of portfolio firm directors attended an institution that is one of the five most common institutions for their respective samples.

2.2 Information transfer from portfolio firm to hedge fund company

We examine how information may be disseminated from portfolio firm to hedge fund company through a shared educational connection between corporate executives at the portfolio firm with their counterparts at the hedge fund company.

2.2.1 Options holdings

We begin by examining the direction of the positions taken by hedge fund companies on portfolio firms, and investigate whether the positions are related to connections between portfolio firms and hedge fund companies. Following the procedure in Aragon and Martin (2012), we use manually collected data on the hedge fund companies' portfolio holdings from the 13-F filings on SEC EDGAR as a proxy for the direction that hedge fund companies expect future returns to move. In contrast to the Thomson Reuters 13-F database, these filings include information on long option positions held by hedge fund companies, which allows us to define the direction of the holdings more precisely. Specifically, hedge fund companies are required to report the number and value of the underlying stock for their long options, as well as the option type (put or call), but are not required to report the strike price or the maturity date of the long options. Since generally only equity securities that trade on an exchange, equity options and warrants, shares of closed-end investment companies, and certain convertible debt securities are considered by the SEC to be official Section 13-F securities, we do not observe other types of derivatives, such as Total Return Swaps. Since options that are written by the reporting institution should not be reported,

we do not observe short option positions as well.

Table 2.2 Panel A describes the summary statistics of the option holdings for a subset of hedge funds companies in our sample. The sample includes 259 unique hedge fund companies that hold options in any given year, and an average hedge fund company holds about 35 such positions per year. The total underlying value of the stock covered by the options per year across all hedge fund companies is around $200 million, and is biased towards the put options. Panel B displays the number of portfolio firms that the hedge fund companies hold options on, as well as the educational connection variables for those portfolio firms. On average, hedge fund companies hold close to 1,137 unique stocks each year and about 450 of these stocks feature an institution-only education connection between a hedge fund company manager and a key portfolio firm personnel. Applying stricter criterion for the educational connection variable reduces the number of connected firms to about 23 per year for educational connections that have the same school, degree type, and graduation year (*Connection3*). We have omitted the educational connection variable (*Connection4*) that includes the criterion that the hedge fund manager be a principal since there are no observations for that variable.

For each hedge fund company and portfolio firm pair, we identify that portfolio holding as a net call or net put position based on the following criteria used in Aragon and Martin (2012). A portfolio holding is defined as a net put position if the amount of underlying stock on the long put options exceeds that of the sum of the long call options and the long equity position. Conversely, a portfolio holding is defined as a net call position if the amount of underlying stock on the long call options exceeds that of the long put options. Note, however, that the latter is not the direct reverse of the former since there is no available data on short equity positions. In the absence of such data, we assume that there are no short equity positions. However, while the data may be incomplete with regards to the short equity positions, we expect that the hedge funds would hold put options for portfolio firms from which they expect negative returns rather than short equity positions. Additionally, since data on the strike price or maturity date is unavailable, we are unable to value weight

the options in classifying the positions.

We then take the portfolio holdings of each hedge fund company and place them in either a net put portfolio or a net a call portfolio based on the identified net put and net call positions. The portfolio holdings are assumed to be fixed over the quarter, and we update the portfolio holdings at the end of every quarter. The monthly returns for the net call and net put portfolios of each hedge fund company are then calculated by aggregating the monthly returns on an equally weighted basis. The monthly returns for the net call and net put portfolios are then aggregated across all the hedge fund companies for every month in our sample, such that we end up with an aggregated net call portfolio and an aggregated net put portfolio. We then use these aggregated portfolios to create a zero-investment long-short portfolio that is (i) long the aggregated net call portfolio and (ii) short the aggregated net put portfolio.

Next, we examine portfolio firms that are connected through *Connection1*, which is if there is at least one pair of hedge fund manager and key portfolio firm personnel that graduated from the same educational institution. We repeat the portfolio building procedure for these firms, such that we have the following three portfolios: first, a portfolio of connected firms that the hedge fund has a net put position in; second, a portfolio of connected firms that the hedge fund has a net call position in; and third, a zero-investment long-short portfolio that is long the net call portfolio and short the net put portfolio. We create these three portfolios for each of the connection variables, as well as unconnected firms, which we define as portfolio firms that do not have fulfill the criteria for any of the connection variables.

Table 2.3 presents the results for the net put, net call, and long-short portfolios. We report the 4-factor alphas, raw returns, and DGTW (Daniel, Grinblatt, Titman, and Wermers, 1997) benchmark-adjusted returns. The net put portfolios exhibit significant negative abnormal returns, while the net call portfolios are largely associated with insignificant abnormal returns. The long-short portfolio abnormal returns are also generally significant and positive. Overall, the abnormal returns increase monotonically in magnitude with the strength of the connection. In particular, the negative returns of the net put portfolio increase in size

from 47 basis points per month for unconnected portfolios to 109 basis points for educational connections that have the same school, degree, and graduation year. The long short portfolios earn an abnormal return of 129 to 137 basis points per month. The negative abnormal returns for the net put portfolios suggest that hedge fund companies expected poor stock performance for the stocks in the portfolio, and have prepared for this eventuality by holding put options. Unfortunately, we do not see evidence of positive abnormal returns for the net call portfolios, but this may be due to lack of available data on short equity positions.

2.2.2 Options around mergers and acquisitions

Since the merger or acquisition is a time sensitive private decision by the acquirer firm, it is more likely that the informational advantage of hedge fund companies lies in the actual informational transfer of event specific information from key portfolio firm personnel to connected hedge fund managers rather than simply due to hedge fund managers being familiar with the business decision making approach of portfolio firm personnel through their shared educational background. If hedge fund companies are able to obtain information through their connections, hedge fund companies would be in a better position to anticipate potential merger or acquisition deals by their portfolio firms, as well as the value outcome of those deals. As such, if educational connections provide an informational advantage to hedge funds, we would expect hedge funds to hold a net call (put) position on acquirers whose deals are likely to exhibit positive (negative) merger or acquisition announcement returns.

We utilise a sub-sample of portfolio firms that engaged in a merger or acquisition deal as an acquirer during our sample period from our main sample. We do not examine portfolio firms that were acquisition targets since these firms are (i) generally smaller and (ii) are less likely to have associated options in the market. We limit the sample to deals where the deal value is at least 1% of the value of the total assets of the acquirer so as to exclude deals with targets that are very small relative to their acquirers. We further limit our sample of portfolio firms to those in which the hedge fund company has option holdings of at the 13-F quarterly filing date closest to and at least one calendar month before the merger announcement date.

When the filing of the 13-F form is delayed, we use the effective date of the Form 13-F in order to ensure that the data on the Form 13-F is the most appropriate with regards to the time.

Table 2.4 reports the summary statistics for the merger and acquisition deals for portfolio firms in our sample that have engaged in such a deal as an acquirer. Our sample consists of 4,557 merger and acquisition deals that are worth on aggregate $2.3 trillion. In contrast, there are 1,288 deals worth $1.5 trillion in which the hedge fund company held options on the acquirer prior to the acquisition. While only 26.9% of all deals involve an acquirer that a hedge fund companies holds options on prior to the deal, the value of the deals on which options are held make up about 66.2% of the total value of all deals. This suggests that hedge funds use options as an alternative investment mechanism on large deals, or invest relatively more on such deals due to a potential informational advantage.

We calculate the acquirer abnormal three day (t-1 to t+1) merger or acquisition announcement returns based on the market model. We define the Announcement Return Direction as 1 if the acquirer's merger or acquisition announcement return is positive, and 0 if it is negative. We also define additional option indicator variables, Put and Call: Put (Call) is defined as equal to 1 if the underlying shares of the put options held by the hedge fund company are more (less) than the underlying shares of the call options held by the hedge fund company, and zero otherwise.

We estimate a logistic regression of the Announcement Return Direction on the connection variables, the option indicator, and the interaction terms between the connection variables and the option indicator. We include the market value of equity and market-to-book ratio of the acquirer, as well as a relative size measure of the acquirer and target to control for size effects. We also include indicator variables for cash only deals and tender offer deals, as the difference in compensation may also denote the quality of the deal. Finally, we include an indicator variable for whether the acquirer and target are in the same industry (based on the 2-digit SIC code) to control for within-industry deals.

The odds ratios for the logistic regressions are reported in Table 2.5. Panel A displays

the results for stocks in which the hedge fund companies hold a net put option position. The likelihood of the acquirer's merger or acquisition announcement returns being positive is statistically unrelated to the presence of a net put option holding by the hedge fund company. The effect of the presence of a net put option on the odds ratio is not large (close to 1), but the interaction term between the stronger educational connection variables and the net put indicator variable report odds ratios that are statistically significant and below one. The results show that when a hedge fund company holds a net put position in a connected portfolio firm, the odds of the merger or acquisition announcement return being positive is 29.5% to 46.4% of the odds when the hedge fund company holds a net put position in a portfolio firm, connected or unconnected. On the other hand, the results for the call options (Panel B) indicate the opposite result. In this instance, the acquirer's merger or acquisition announcement return is 2.0 times to 3.4 times more likely to be positive when the hedge fund company holds a net call option position in a connected acquirer compared to if the hedge fund company simply has a net call position in the acquirer, connected or unconnected. However, this result only holds for the stronger educational connection variables.

Overall, the results presented in Table 2.5 suggest that hedge funds can predict the direction of a connected acquirer's merger or acquisition announcement returns. The net call (put) positions taken by the hedge fund company in connected acquirers are positively (negatively) related to the likelihood that the acquirer's merger or acquisition announcement returns is positive compared to when the hedge fund company simply holds net call (put) positions in any acquirer, connected and unconnected. Further, the results also suggest that the ability of the hedge fund companies to predict the direction of the acquirer's merger or acquisition announcement returns only manifests when the hedge fund company has a strong educational connection to the directors or managers of the acquirer. Our findings suggest that event-specific information transfer for mergers and acquisitions undertaken by portfolio firms only occur for strong educational connections between the hedge fund company and portfolio firms, and that the information being transmitted has market value.

However, the preceding results do not eliminate the possibility that the preceding results

are due to a long-term investment strategy, rather than informed opportunistic trading by the hedge fund companies. In order to distinguish the two, we modify and repeat the analysis by using 13-F filings made one quarter before and after the 13-F filings that were used in the regressions for Table 2.5 to construct the positional portfolios. If the results of our prior analysis are due to a long-term strategy, then we would expect the results based on these other 13-F filings made one quarter before and after to be similar to that of the prior analysis. However, we find limited evidence that the net option positions taken by hedge funds at the later and earlier filing dates are related to the merger or acquisition announcement return (in Tables 2.6 and 2.7 respectively). The net option position taken just prior to the merger is therefore likely to be a transitory and opportunistic investment decision made by the hedge fund companies in our sample, possibly informed by the educational connection between the hedge fund companies and their portfolio firms.

2.3 Hedge fund company portfolios

One potential explanation for our results in the preceding section is that hedge fund companies invest differently compared to other investors examined in the extant literature. We examine our sample in the same manner as Cohen, Frazzini, and Malloy (2008) by establishing the choices that the hedge funds make with regards to the amount that they invest in connected firms relative to unconnected firms, as well as the performance of their investments in connected firms relative to unconnected firms. We establish that the investment behavior of the hedge fund companies in our sample with regards to the educational connection variables are similar to those of the mutual fund companies examined by Cohen, Frazzini, and Malloy (2008), and therefore confirm that our results are unlikely to be due to differences in the investor type being evaluated.

2.3.1 Portfolio weights

We begin by examining hedge fund company portfolio weights, which are defined as the fraction of the value of the stock holdings over the total value of all the equity holdings of

the hedge fund company. As the fund-firm educational connection is the variable of interest, we implement regressions of the portfolio weights on the connection variables that we have defined previously, following the regression structure used by Cohen, Frazzini, and Malloy (2008).

Table 2.8 Panel A reports that the coefficients for the connection indicator variables are all positive and highly significant. We find that hedge fund companies assign a portfolio weight that is 2.62 basis points larger for a portfolio firm that has an educational connection than a portfolio firm without such a connection. The magnitude of the over-weighting monotonically increases with the strength of the educational connection, as a portfolio firm that has an educational connection that is based on the same school, graduation year, and degree type (*Connection4*) has a portfolio weight that is 16.91 basis points more on average than a portfolio firm that does not have such a connection. Restricting the educational connection to hedge fund principals only further increases the over-weighting to 40.10 basis points. These results are robust to stock, industry, time and hedge fund company fixed effects (Panel B), and consistent with those of Cohen, Frazzini, and Malloy (2008).

Next, we examine the sensitivity of the results to the graduation year restriction. We modify the two educational connection variables *Connection3* and *Connection4* by allowing the graduation years to differ by one or two years, instead of restricting it to being the same year. This relaxed restriction allows us to pick up relationships between people who were plausibly at the same school at the same time, rather than relationships between people who graduated in the same cohort. Our results (Panel C) show that relaxing the time criterion reduces the magnitude of the over-weighting by about half (16.91 basis points as compared to 8.43 and 8.00 basis points), indicating that while there is an effect from an educational connection based on being at the same school at the same time, it is still weaker than the effect from the educational connection based on graduating in the same year.

Finally, in Panel D, we further modify the educational connection variable by splitting both *Connection3* and *Connection4* into two based on whether the educationally connected portfolio firm directors are senior or junior directors. We define senior directors as directors

who are board committee chairs or the chair of the board, while junior directors are defined as all other directors. We find that educational connections with senior directors are associated with portfolio over-weighting that is twice as large as that with junior directors (21.40 basis points versus 9.69 basis points). Similarly, educational connections between senior directors and senior hedge fund principals more than doubles the magnitude of over-weighting (48.49 basis points versus 21.40 basis points). Assuming that the senior directors have access to more information and the senior hedge fund principals have greater say over investment decisions, the differences in the results across seniority suggest that there is greater information transfer and usage between educationally connected senior hedge fund principals and portfolio firm directors.

2.3.2 Portfolio returns

We test for a difference in the portfolio returns on equity holdings across connected and unconnected firms in our sample of hedge fund companies and find similar results to those found by Cohen, Frazzini, and Malloy (2008) on mutual fund companies.

We assign the stocks held by each hedge fund company to two portfolios (connected and unconnected) for each hedge fund company, on the basis of whether the portfolio firm associated with the stock has an educational connection with the hedge fund company. We then calculate the calendar-time monthly weighted average return for the two portfolios of each hedge fund company, assuming that the hedge funds do not modify their portfolios until the portfolio weights are updated at the end of each quarter since the 13-F is reported on a quarterly basis. The monthly weighted average return of each portfolio is then aggregated across all of the hedge fund companies, weighted on the fraction of each hedge fund companys assets under management over the total assets under management of all the hedge fund companies in the sample. Finally, we compute monthly long-short aggregated portfolio returns by subtracting the unconnected portfolio returns from the connected portfolio returns. We repeat the process by instead forming the connected and unconnected portfolios based on the various connection variables that we have defined previously.

Table 2.9 Panel A displays the 4-factor alphas based on the Carhart (1997) model, raw returns, and DGTW-benchmark-adjusted returns for connections of differing strength. Overall, the educationally connected portfolios achieve a statistically significant higher return compared to the educationally unconnected portfolios. Additionally, the effect is monotonically increasing in the strength of the educational connection for the long-short portfolios. The long-short portfolio strategy earns a monthly 4-factor alpha of up to 80 basis points and a monthly DGTW-benchmark-adjusted monthly alpha of up to 48 basis points. In contrast, the monthly alphas for the unconnected portfolios are largely insignificant. Additionally, the portfolios that are educationally connected through the weakest connection level do not exhibit any significant alpha, denoting that weaker educational connections are less likely to be related to returns.

As an additional check, we also contrast the performance of stocks that are educationally connected to but not held by the hedge fund companies with the stocks that are educationally connected to and held by the hedge fund companies (Panel B). We implement a similar procedure to the one used to compare the educationally connected and unconnected stocks in the hedge fund portfolios. We begin with a set of all firms in CRSP whose directors and managers share a connection with each specific hedge fund company, and then assign them to portfolios based on whether the stock is held by the hedge fund company. We find similar results to those previously shown in Panel A: the educationally connected portfolios achieve a statistically significant higher return that is monotonically increasing in connection strength. The strategy that is long the educationally connected-and-held stocks and short the educationally connected-and-not-held stocks earns a monthly 4-factor alpha of up to 72 basis points, and a monthly DGTW-benchmark-adjusted monthly alpha of up to 48 basis points. These results suggest that the portfolio over-weighting towards connected firms may be motivated by information passing through social networks from firm managers and directors to hedge fund managers, and that the information being transferred is price-critical.

Panel C reports the 4-factor alpha, raw returns, and DGTW-benchmark adjusted returns on the portfolios of educationally connected stocks only. Panel C-1 shows the results for

educational connections involving senior directors at portfolio firms, whereas Panel C-2 shows the results for educational connections involving junior directors at portfolio firms. Similarly, Table 2.9 Panel C shows that stock holdings that are educationally connected through a senior portfolio firm director outperform stock holdings that are educationally connected through a junior portfolio firm director.

Table 2.9 Panel D reports the connected portfolio returns when observations of hedge fund company and portfolio firm pairs located in the same state is excluded, in order to control for the possibility of home bias in the investment decisions of the hedge fund companies in the sample. We find that the initial results still hold even after controlling for the potential home-bias.

Overall, the results in this section show that the hedge fund companies in our sample behave similarly to the mutual fund companies examined in Cohen, Frazzini, and Malloy (2008). As such, it is unlikely that our results are driven by the hedge fund companies we examine. Additionally, it suggests that our results are generalisable and applicable to all investors.

2.4 Conclusion

We investigate a potential information channel through which private corporate information can be disseminated to the market. Specifically, we examine social networks between hedge fund companies and their portfolio firms based on shared educational background and history in order to answer the challenge posed by Cohen, Frazzini, and Malloy (2008) to disentangle whether socially connected investments perform better due to information transfer through the connection or an informational advantage based on a common background.

We find that a long-short strategy based on educationally connected option holdings earns positive abnormal returns. We also examine portfolio firms that are involved in a merger or acquisition as an acquirer, and find that a net call (put) option position held by hedge fund companies on these acquirers shortly before the merger or acquisition announcement date are associated with a higher (lower) likelihood that the merger or acquisition announcement

returns are positive. Given that the decision to undertake the deal is a time-critical private decision of the firm, it is more likely that the ability of the hedge fund company in predicting the direction of the merger or acquisition announcement returns of educationally connected acquirers in their portfolio is due to information transfer from the portfolio firm to hedge fund company rather than an informational advantage due to a familiarity with the approach or preferences of the portfolio firm executives based on a common educational history.

We also establish that hedge fund companies perform similarly to the mutual fund companies examined in Cohen, Frazzini, and Malloy (2008), and therefore confirm that our results are not due to the investor type being evaluated. Hedge fund companies over-weight portfolios in favor of connected firms, and earn abnormal returns on these holdings of connected firms. Buy-and-hold portfolios of hedge fund holdings based on publicly available information generate abnormal returns even after accounting for traditional risk factors. Our results support the hypothesis that hedge fund companies obtain and use information from their firm-side connections to inform their investment decisions.

Table 2.1: Summary statistics

This table presents summary statistics of the hedge fund firm sample. We match hedge fund firms in the Barclays Hedge Fund database with the quarterly institutional 13-F holdings data from Thomson Reuters. Only the matched firms whose primary business is hedge fund management are included in the sample. Panel A reports the annual average number of the hedge fund firms, the number of long stock holdings, the average and median dollar amounts of the holdings in the quarterly 13-F data. The quarterly 13-F filings are from March 2007 to March 2015. In Panel B, for each year, we report the total number of stocks owned by the sample hedge fund firms. Many of these stocks may be owned by multiple firms. Column *Connection1* reports the number of stocks whose board members attended the same universities as the managers at any sample hedge fund firms that own the stocks. Column *Connection2A* reports the number of stocks whose board members received the same degrees at the same universities as the sample fund managers. Column *Connection2B* reports the number of stocks whose board members received degrees in the same years at the same universities as the sample fund managers. Column *Connection3* reports the number of stocks whose board members received the same degrees at the same universities in the same years as the sample fund managers. Column *Connection4* reports the number of stocks whose board members received the same degrees at the same universities in the same years as the principal managers of the sample hedge fund firms. Panel C reports the sample statistics of fund managers in the sample hedge fund firms and directors in the publicly traded sample firms as well as the university degrees they earned. We manually collect the educational information of the hedge fund managers from multiple sources. We obtain the directors educational data from the BoardEx database. Panel D shows the distribution of the graduation years of the sample hedge fund managers and of the directors of the sample public firms. Panel E reports the top five most attended universities and the fractions of the degrees earned from the top universities to the total degrees earned by each group in the sample.

Table 2.1: Summary statistics

Panel A: Hedge fund manager characteristics

Year	Hedge fund firms (N = 429)	Average number of stocks	Mean ($ million)	Median ($ million)
2007	275	89	784	254
2008	286	79	574	149
2009	259	81	530	127
2010	243	89	706	210
2011	255	83	695	198
2012	250	81	696	192
2013	255	86	846	241
2014	275	85	910	239
2015Q1	272	82	926	238
Average	262	84	724	205

Panel B: Number of unique stocks by connection

Year	Stocks	Connection1	Connection2A	Connection2B	Connection3	Connection4
2007	4,167	2,613	2,321	485	207	127
2008	4,013	2,531	2,237	459	197	115
2009	3,743	2,341	2,074	413	199	100
2010	3,546	2,217	1,997	401	173	91
2011	3,453	2,279	2,011	422	208	104
2012	3,444	2,333	2,041	405	190	101
2013	3,441	2,362	2,120	487	232	109
2014	3,256	2,224	1,993	503	246	87
2015Q1	2,819	1,818	1,591	310	119	33

Panel C: Total and average manager count

Entity	Type	Total count	Average per firm-year	Median per firm-year
Hedge fund firms	Individuals	5,240	7	5
	Degrees	8,095	11	7
Portfolio firms	Individuals	25,718	5	5
	Degrees	41,530	9	8

Panel D: Breakdown of graduation years

Degree Year	Hedge fund managers	Public firm directors
Before 1970	2%	38%
1970-1979	4%	34%
1980-1989	15%	20%
1990-1999	32%	6%
After 1999	48%	1%
	100%	100%

Panel E: Top five most attended universities

Rank	Hedge fund managers	%	Portfolio firm directors	%
1	Univ. Pennsylvania	6%	Harvard University	7%
2	Columbia University	5%	Stanford University	3%
3	Harvard University	4%	Univ. Pennsylvania	3%
4	New York University	4%	Columbia University	2%
5	Univ. Chicago	3%	Yale University	2%

Table 2.2: Sample statistics of option holdings

This table presents the sample statistics of options holdings by the sample hedge fund firms. The information on the hedge fund option holdings was manually collected from the fund companies original 13-F filings. Panel A reports the total number of the option holders among the sample hedge fund firms, the average number of positions held and the value of the underlying stocks per each firm-year. The panel also shows a breakdown of the number of the holdings and the value of the underlying stocks by option type. In Panel B, for each year, we report the total number of underlying stocks of which the sample hedge fund firms hold option positions. Multiple hedge fund firms may hold options on the same underlying stocks. The definitions of the connections in the panel are the same as in Table 2.1.

Table 2.2: Sample statistics of option holdings

Panel A: Count of hedge fund managers with single stock options per firm-year

Year	HF option holders (N = 259)	Average option positions	Underlying stock value ($1,000)	Call positions	Underlying stock value ($1,000)	Put positions	Underlying stock value ($1,000)
2007	104	31	131,229	16	68,998	15	62,231
2008	109	32	94,274	16	48,509	15	45,766
2009	84	43	146,033	23	81,610	19	64,422
2010	92	31	170,150	16	93,660	15	76,490
2011	96	32	202,991	16	108,625	16	94,366
2012	87	40	211,068	22	105,532	19	105,535
2013	96	35	258,178	18	147,100	16	111,078
2014	109	27	227,385	15	153,401	12	73,984
2015Q1	101	23	197,938	12	132,056	11	65,881

Panel B: Number of unique underlying stocks by connection

Year	All option stocks	Connected1	Connected2A	Connected2B	Connected3
2007	1,100	462	378	53	17
2008	1,280	507	425	67	28
2009	1,229	469	361	43	19
2010	1,134	361	293	39	17
2011	1,154	465	345	51	27
2012	1,145	491	381	54	30
2013	1,151	487	363	62	33
2014	984	383	264	24	13
2015Q1	1,055	427	328	47	26

Table 2.3: Hedge fund manager connections and option performance

This table presents calendar time portfolio returns for the 2007-2015 1Q sample period. We collect information on the hedge fund firms single-stock options holdings from the original 13-F filings obtained from SEC EDGAR. A net call is a call position where the number of shares of underlying stock on the call position exceeds that of a put position held by the same hedge fund firm on the same stock. A net put is a put position where the number of shares of underlying stock on the put position exceeds that of a call position and of a long equity position held by the same hedge fund firm on the same stock. In Panel A, we assign each underlying stock of the options held by each sample hedge fund firm to either the Net Call or Net Put portfolios at the beginning of each quarter on the basis of the option holding type. We calculate the monthly equal-weighted average return of each portfolio, assuming that the hedge fund firms maintain their option holdings until the holdings are updated at the end of each quarter. We then compute the monthly equal-weighted average return of each portfolio across the sample hedge fund firms. Panel A display the intercepts (alpha) from the regressions of the monthly excess portfolio returns on the Fama and French (1993) three factors and the Carhart (1997) momentum factor; the monthly portfolio returns computed using the raw stock returns, averaged over the sample period; and the monthly portfolio returns computed using the Daniel, Grinblatt, Titman, and Wermers (1997) benchmark-adjusted stock returns, averaged over the sample period. In Panel B, we assign each underlying stock of the options held by each sample hedge fund firm to the Non-directional portfolio at the beginning of each quarter, which comprises of the remaining option positions that are not assigned to either the Net call or Net put portfolios. ***, **, * indicate significance at 1%, 5% and 10% levels, respectively.

Table 2.3: Hedge fund manager connections and option performance

Panel A: Directional option performance

	4 Factor alpha			Raw return			Benchmark-adjusted return		
Connection	**Net Call**	**Net Put**	**Long-short**	**Net Call**	**Net Put**	**Long-short**	**Net Call**	**Net Put**	**Long-short**
Unconnected	-0.0033**	-0.0047***	0.0014	0.0060	0.0044	0.0016	-0.0012	-0.0023	0.0011
	(2.5486)	(3.2915)	(1.5063)	(0.9419)	(0.6450)	(1.3842)	(-0.9834)	(-1.3316)	(1.1321)
Connection1	-0.0013	-0.0049***	0.0036***	0.0074	0.0035	0.0039**	0.0002	-0.0035*	0.0037**
	(1.2497)	(3.8003)	(2.6750)	(1.1642)	(0.4958)	(2.4338)	(0.1708)	(-1.8291)	(2.4700)
Connection2A	-0.0009	-0.0054***	0.0046***	0.0079	0.0030	0.0048***	0.0005	-0.0040**	0.0045***
	(0.6719)	(3.9803)	(3.3198)	(1.1997)	(0.4218)	(3.0520)	(0.3863)	(-2.0280)	(3.0468)
Connection2B	0.0014	-0.0090***	0.0104***	0.0099	-0.0007	0.0106***	0.0023	-0.0078**	0.0100***
	(0.5934)	(3.0387)	(2.6869)	(1.5385)	(-0.1035)	(2.8091)	(1.0138)	(-2.6134)	(2.7784)
Connection3	0.0027	-0.0109**	0.0137**	0.0112	-0.0017	0.0129**	0.0037	-0.0093*	0.0130**
	(0.8854)	(2.1608)	(2.3707)	(1.6544)	(-0.2082)	(2.2688)	(1.3696)	(-1.9150)	(2.5058)

Panel B. Non-directional option performance

Connection type	**4 Factor alpha**	**Raw return**	**Benchmark-adj. return**
Unconnected	-0.0047***	0.0045	-0.0023
	(2.9768)	(0.7034)	(-1.6084)
Connection1	-0.0021*	0.0067	-0.0006
	(1.7833)	(1.0406)	(-0.4077)
Connection2A	-0.0022	0.0066	-0.0005
	(1.5638)	(0.9709)	(-0.2936)
Connection2B	-0.0036	0.0040	-0.0035
	(0.7136)	(0.4963)	(-0.7559)
Connection3	-0.0012	0.0054	-0.0013
	(0.1462)	(0.4908)	(-0.1737)

Table 2.4: M&As with options held on acquirers in month $t-1$

This table reports the descriptive statistics of the sample M&A deals. Only the deals by public acquirers on public, private, or subsidiaries targets with the ratio of the deal value to the market value of acquirer assets greater than 2% are included in the M&A sample. The M&A sample is from the Thomson Reuters SDC database. Columns Number of M&As and the Deal value show the total number and the dollar value of all M&A deals per year. Columns M&A options and Deal value with options show the number and the dollar value of the M&A deals where the sample hedge fund firms hold options on the acquirer's stocks one month prior to merger or acquisition announcement.

Table 2.4: M&As with options held on acquirers in month $t-1$

Year	Number of M&As	M&A options	Deal value ($ millions)	Deal value with options ($ millions)
2007	792	194	330,546	156,826
2008	639	189	264,570	193,915
2009	453	181	381,384	318,487
2010	501	152	236,006	156,573
2011	516	132	270,518	198,349
2012	559	162	239,104	160,284
2013	432	110	214,494	121,871
2014	539	136	242,161	145,584
2015Q1	126	32	127,901	76,199
Total	4,557	1,288	2,306,686	1,528,088

Table 2.5: Hedge fund option holdings and merger or acquisition announcement returns

This table presents the logistic regression results of merger or acquisition announcement return predictability for the 2007-2015Q1 sample period. The dependent variable is a binary variable that takes the value 1 if an acquirers merger or acquisition announcement return is positive and take 0 if it is negative. The announcement return is measured using the three-day cumulative abnormal return based on the market model (Brown and Warner 1985) during the $[-1, +1]$ event window surrounding the announcement day. Panel A reports the logistic regression results of the binary dependent variable on the put positions held by the sample hedge fund firms. *Put option* is a binary variable that takes the value 1 if a hedge fund firm holds a net put position three month before the month when a merger or acquisition announcement takes place and 0 otherwise. *Market equity* is the market value of an acquirer one month prior to the merger or acquisition announcement month. *MB assets* is (market equity + short term debt + long term debt + liquidation value of preferred stock)/Book value of total assets of an acquirer. *Cash only* is a binary variable that takes the value 1 if a merger is of a 100% cash deal and 0 otherwise, while *Stock only* is a binary variable that takes the value 1 if a merger is of a 100% stock deal and 0 otherwise. *Tender offer* is a binary variable that takes the value 1 if the deal is a tender offer and 0 otherwise. *Concentrated* is a binary variable that takes the value 1 if the acquirer and the target belong to the same 2-digit SIC code and 0 otherwise. *Relative size* is the ratio of the dollar value of the deal to the total market value of the acquirers assets, measured one month prior to the announcement month. Panel B reports the logistic regression results of the binary dependent variable on the call positions held by the sample hedge fund firms. Call option is a binary variable that takes the value 1 if a hedge fund firm holds a net call position one month prior to the month when a merger or acquisition announcement takes place. z-statistics from standard errors clustered by quarter is reported in the parentheses. ***, **, * indicate significance at 1%, 5% and 10% levels, respectively.

Table 2.5: Hedge fund option holdings and merger or acquisition announcement returns

Panel A: Put option holdings and acquirer's merger or acquisition announcement returns

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Net Put× Connection1	1.017				
	(0.121)				
Net Put× Connection2A		1.129			
		(0.762)			
Net Put× Connection2B			0.464***		
			(-2.154)		
Net Put× Connection3				0.361**	
				(-2.021)	
Net Put× Connection4					0.295*
					(-1.907)
Net Put option	1.056	1.028	1.088	1.081	1.074
	(0.567)	(0.291)	(1.004)	(0.934)	(0.859)
Connection1	1.026				
	(0.266)				
Connection2A		0.992			
		(-0.075)			
Connection2B			1.662*		
			(1.947)		
Connection3				2.243**	
				(2.027)	
Connection4					3.025**
					(2.132)
Market equity	0.908**	0.909**	0.909**	0.908**	0.908**
	(-2.270)	(-2.236)	(-2.254)	(-2.270)	(-2.285)
MB Assets	0.856*	0.856*	0.854*	0.854*	0.854*
	(-1.836)	(-1.842)	(-1.867)	(-1.864)	(-1.864)
Cash only	1.024	1.025	1.023	1.020	1.019
	(0.153)	(0.159)	(0.150)	(0.131)	(0.124)
Stock only	1.085	1.087	1.078	1.084	1.082
	(0.243)	(0.251)	(0.224)	(0.242)	(0.237)
Tender offer	1.001	1.003	1.003	1.004	1.002
	(0.00223)	(0.00919)	(0.0116)	(0.0149)	(0.00873)
Concentrated	1.153	1.154	1.153	1.150	1.149
	(1.060)	(1.070)	(1.064)	(1.048)	(1.038)
Relative size	1.313	1.310	1.314	1.314	1.312
	(0.807)	(0.801)	(0.811)	(0.809)	(0.806)
Constant	3.516*	3.498*	3.460*	3.511*	3.558*
	(1.867)	(1.853)	(1.834)	(1.861)	(1.890)
Observations	3,723	3,723	3,723	3,723	3,723
Pseudo R-squared	0.0222	0.0222	0.0222	0.0222	0.0222

Panel B: Call option holdings and acquirer's merger or acquisition announcement returns

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Net Call × Connection1	0.893				
	(-0.861)				
Net Call × Connection2A		0.760*			
		(-1.864)			
Net Call × Connection2B			2.065**		
			(2.076)		
Net Call × Connection3				2.784**	
				(1.963)	
Net Call × Connection4					3.493*
					(1.901)
Net Call option	0.983	1.013	0.917	0.922	0.927
	(-0.180)	(0.138)	(-1.026)	(-0.957)	(-0.890)
Connection1	1.091				
	(0.985)				
Connection2A		1.190			
		(1.517)			
Connection2B			0.859		
			(-0.687)		
Connection3				0.880	
				(-0.424)	
Connection4					0.973
					(-0.0723)
Market equity	0.908**	0.910**	0.907**	0.907**	0.907**
	(-2.257)	(-2.218)	(-2.288)	(-2.295)	(-2.311)
MB Assets	0.856*	0.856*	0.854*	0.854*	0.854*
	(-1.833)	(-1.837)	(-1.863)	(-1.866)	(-1.868)
Cash only	1.023	1.024	1.023	1.022	1.020
	(0.148)	(0.152)	(0.146)	(0.146)	(0.133)
Stock only	1.088	1.089	1.085	1.089	1.087
	(0.254)	(0.256)	(0.244)	(0.257)	(0.250)
Tender offer	1.001	1.007	1.000	1.000	0.999
	(0.00247)	(0.0262)	(-0.00136)	(0.000402)	(-0.00333)
Concentrated	1.156	1.156	1.153	1.151	1.149
	(1.081)	(1.084)	(1.061)	(1.047)	(1.038)
Relative size	1.316	1.317	1.311	1.314	1.310
	(0.812)	(0.816)	(0.804)	(0.809)	(0.802)
Constant	3.587*	3.464*	3.827**	3.833**	3.870**
	(1.896)	(1.853)	(1.993)	(2.000)	(2.023)
Observations	3,723	3,723	3,723	3,723	3,723
Pseudo R-squared	0.0222	0.0222	0.0222	0.0222	0.0222

Table 2.6: Post-announcement option holdings and announcement returns

This table presents the logistic regression results of merger or acquisition announcement return predictability for the 2007-2015Q1 sample period. The dependent variable is a binary variable that takes the value 1 if an acquirers merger or acquisition announcement return is positive and take 0 if it is negative. The announcement return is measured using the three-day cumulative abnormal return based on the market model (Brown and Warner 1985) during the $[-1, +1]$ event window surrounding the announcement day. Panel A reports the logistic regression results of the binary dependent variable on the put positions held by the sample hedge fund firms. *Put option* is a binary variable that takes the value 1 if a hedge fund firm holds a net put position three month before the month when a merger or acquisition announcement takes place and 0 otherwise. *Market equity* is the market value of an acquirer one month prior to the merger or acquisition announcement month. *MB assets* is (market equity + short term debt + long term debt + liquidation value of preferred stock)/Book value of total assets of an acquirer. *Cash only* is a binary variable that takes the value 1 if a merger is of a 100% cash deal and 0 otherwise, while *Stock only* is a binary variable that takes the value 1 if a merger is of a 100% stock deal and 0 otherwise. *Tender offer* is a binary variable that takes the value 1 if the deal is a tender offer and 0 otherwise. *Concentrated* is a binary variable that takes the value 1 if the acquirer and the target belong to the same 2-digit SIC code and 0 otherwise. *Relative size* is the ratio of the dollar value of the deal to the total market value of the acquirers assets, measured one month prior to the announcement month. Panel B reports the logistic regression results of the binary dependent variable on the call positions held by the sample hedge fund firms. Call option is a binary variable that takes the value 1 if a hedge fund firm holds a net call position one month prior to the month when a merger or acquisition announcement takes place. z-statistics from standard errors clustered by quarter is reported in the parentheses. ***, **, * indicate significance at 1%, 5% and 10% levels, respectively.

Table 2.6: Post-announcement option holdings and announcement returns

Panel A: Post-announcement put option holdings and acquirers merger or acquisition announcement returns

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Net Put × Connection1	1.071				
	(0.455)				
Net Put × Connection2A		1.051			
		(0.336)			
Net Put × Connection2B			1.062		
			(0.198)		
Net Put × Connection3				0.594	
				(-1.148)	
Net Put × Connection4					0.378
					(-1.350)
Net Put option	1.035	1.047	1.061	1.071	1.070
	(0.352)	(0.529)	(0.799)	(0.922)	(0.900)
Connection1	1.038				
	(0.399)				
Connection2A		1.054			
		(0.548)			
Connection2B			0.864		
			(-0.712)		
Connection3				1.053	
				(0.179)	
Connection4					2.134
					(1.436)
Market equity	0.861***	0.861***	0.864***	0.863***	0.864***
	(-3.620)	(-3.582)	(-3.531)	(-3.550)	(-3.562)
MB Assets	0.930	0.929	0.929	0.929	0.928
	(-0.937)	(-0.944)	(-0.950)	(-0.952)	(-0.958)
Cash only	1.041	1.042	1.044	1.043	1.039
	(0.298)	(0.304)	(0.321)	(0.313)	(0.285)
Stock only	0.780	0.780	0.775	0.776	0.777
	(-0.785)	(-0.787)	(-0.807)	(-0.802)	(-0.796)
Tender offer	1.285	1.287	1.280	1.279	1.284
	(0.934)	(0.941)	(0.921)	(0.917)	(0.936)
Concentrated	1.096	1.095	1.094	1.095	1.096
	(0.700)	(0.693)	(0.684)	(0.693)	(0.695)
Relative size	0.898	0.894	0.894	0.889	0.887
	(-0.340)	(-0.355)	(-0.351)	(-0.370)	(-0.378)
Constant	10.54***	10.49***	10.03***	10.16***	10.13***
	(3.541)	(3.508)	(3.451)	(3.461)	(3.467)
Observations	4,013	4,013	4,013	4,013	4,013
Pseudo R-squared	0.0200	0.0200	0.0200	0.0200	0.0200

Panel B: Post-announcement call option holdings and acquirer's merger or acquisition announcement returns

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Net Call × Connection1	1.103				
	(-0.0783)				
Net Call × Connection2A		1.041			
		(0.746)			
Net Call × Connection2B			0.853		
			(0.284)		
Net Call × Connection3				1.327	
				(-0.509)	
Net Call × Connection4					2.756
					(1.197)
Net Call option	0.932	0.959	0.974	0.965	0.964
	(-0.755)	(-0.512)	(-0.375)	(-0.511)	(-0.522)
Connection1	1.006				
	(0.0622)				
Connection2A		1.049			
		(0.445)			
Connection2B			0.958		
			(-0.167)		
Connection3				0.779	
				(-0.683)	
Connection4					0.938
					(-0.130)
Market equity	0.860***	0.860***	0.864***	0.863***	0.863***
	(-3.652)	(-3.615)	(-3.553)	(-3.568)	(-3.587)
MB Assets	0.931	0.930	0.929	0.929	0.929
	(-0.924)	(-0.940)	(-0.948)	(-0.950)	(-0.957)
Cash only	1.041	1.040	1.043	1.042	1.038
	(0.295)	(0.291)	(0.314)	(0.302)	(0.276)
Stock only	0.780	0.779	0.773	0.774	0.775
	(-0.783)	(-0.788)	(-0.813)	(-0.812)	(-0.805)
Tender offer	1.285	1.288	1.281	1.279	1.284
	(0.935)	(0.943)	(0.928)	(0.919)	(0.935)
Concentrated	1.096	1.094	1.093	1.095	1.095
	(0.695)	(0.686)	(0.676)	(0.688)	(0.690)
Relative size	0.891	0.893	0.897	0.893	0.891
	(-0.364)	(-0.360)	(-0.342)	(-0.358)	(-0.368)
Constant	11.35***	11.12***	10.50***	10.66***	10.70***
	(3.575)	(3.535)	(3.477)	(3.498)	(3.516)
Observations	4,013	4,013	4,013	4,013	4,013
Pseudo R-squared	0.0198	0.0198	0.0198	0.0198	0.0198

Table 2.7: Pre-announcement option holdings and announcement returns

This table presents the logistic regression results of merger or acquisition announcement return predictability for the 2007-2015Q1 sample period. The dependent variable is a binary variable that takes the value 1 if an acquirers merger or acquisition announcement return is positive and take 0 if it is negative. The announcement return is measured using the three-day cumulative abnormal return based on the market model (Brown and Warner 1985) during the $[-1, +1]$ event window surrounding the announcement day. Panel A reports the logistic regression results of the binary dependent variable on the put positions held by the sample hedge fund firms. *Put option* is a binary variable that takes the value 1 if a hedge fund firm holds a net put position three month before the month when a merger or acquisition announcement takes place and 0 otherwise. *Market equity* is the market value of an acquirer one month prior to the merger or acquisition announcement month. *MB assets* is (market equity + short term debt + long term debt + liquidation value of preferred stock)/Book value of total assets of an acquirer. *Cash only* is a binary variable that takes the value 1 if a merger is of a 100% cash deal and 0 otherwise, while *Stock only* is a binary variable that takes the value 1 if a merger is of a 100% stock deal and 0 otherwise. *Tender offer* is a binary variable that takes the value 1 if the deal is a tender offer and 0 otherwise. *Concentrated* is a binary variable that takes the value 1 if the acquirer and the target belong to the same 2-digit SIC code and 0 otherwise. *Relative size* is the ratio of the dollar value of the deal to the total market value of the acquirers assets, measured one month prior to the announcement month. Panel B reports the logistic regression results of the binary dependent variable on the call positions held by the sample hedge fund firms. Call option is a binary variable that takes the value 1 if a hedge fund firm holds a net call position one month prior to the month when a merger or acquisition announcement takes place. z-statistics from standard errors clustered by quarter is reported in the parentheses. ***, **, * indicate significance at 1%, 5% and 10% levels, respectively.

Table 2.7: Pre-announcement option holdings and merger or acquisition announcement returns

Panel A: Pre-announcement put option holdings and acquirer's merger or acquisition announcement returns

Variables	(1)	(2)	(3)	(4)	(5)
Net Put × Connection1	1.043				
	(-1.588)				
Net Put × Connection2A		1.020			
		(-1.028)			
Net Put × Connection2B			0.605		
			(-1.204)		
Net Put × Connection3				0.287**	
				(-2.102)	
Net Put × Connection4					0.396
					(-1.329)
Net Put option	1.143	1.155	1.184**	1.188**	1.175**
	(1.371)	(1.555)	(2.095)	(2.210)	(2.066)
Connection1	1.122				
	(1.249)				
Connection2A		1.123			
		(1.277)			
Connection2B			1.135		
			(0.573)		
Connection3				1.763*	
				(1.794)	
Connection4					1.713
					(1.390)
Market equity	0.906**	0.907**	0.910**	0.909**	0.909**
	(-2.523)	(-2.501)	(-2.414)	(-2.440)	(-2.433)
MB Assets	0.896	0.895	0.890	0.890	0.890
	(-1.372)	(-1.398)	(-1.466)	(-1.459)	(-1.458)
Cash only	1.099	1.101	1.100	1.095	1.099
	(0.595)	(0.609)	(0.603)	(0.571)	(0.593)
Stock only	1.026	1.029	1.019	1.016	1.016
	(0.0792)	(0.0887)	(0.0594)	(0.0483)	(0.0501)
Tender offer	1.141	1.146	1.143	1.147	1.143
	(0.519)	(0.535)	(0.523)	(0.538)	(0.524)
Concentrated	1.199	1.195	1.195	1.195	1.195
	(1.427)	(1.402)	(1.401)	(1.399)	(1.401)
Relative size	1.337	1.322	1.305	1.302	1.309
	(0.780)	(0.754)	(0.713)	(0.708)	(0.723)
Constant	3.886**	3.874**	3.744**	3.789**	3.781**
	(2.070)	(2.058)	(2.004)	(2.023)	(2.026)
Observations	3,681	3,681	3,681	3,681	3,681
Pseudo R-squared	0.0183	0.0183	0.0183	0.0183	0.0183

Panel B: Pre-announcement call option holdings and acquirer's merger or acquisition announcement returns

Net Call × Connection1	1.070				
	(0.565)				
Net Call × Connection2A		0.977			
		(-0.177)			
Net Call × Connection2B			1.715		
			(1.428)		
Net Call × Connection3				2.695*	
				(1.825)	
Net Call × Connection4					2.383
					(1.331)
Net Call option	0.853*	0.883	0.857*	0.859*	0.866*
	(-1.772)	(-1.449)	(-1.889)	(-1.926)	(-1.833)
Connection1	1.090				
	(0.860)				
Connection2A		1.139			
		(1.269)			
Connection2B			0.722		
			(-1.137)		
Connection3				0.698	
				(-0.964)	
Connection4					0.814
					(-0.468)
Market equity	0.904**	0.905**	0.908**	0.907**	0.907**
	(-2.565)	(-2.536)	(-2.472)	(-2.496)	(-2.489)
MB Assets	0.896	0.894	0.890	0.890	0.890
	(-1.377)	(-1.396)	(-1.461)	(-1.461)	(-1.461)
Cash only	1.097	1.099	1.097	1.095	1.098
	(0.587)	(0.599)	(0.587)	(0.571)	(0.590)
Stock only	1.030	1.028	1.025	1.023	1.021
	(0.0932)	(0.0874)	(0.0776)	(0.0696)	(0.0658)
Tender offer	1.136	1.144	1.139	1.144	1.141
	(0.502)	(0.526)	(0.508)	(0.527)	(0.518)
Concentrated	1.200	1.198	1.199	1.196	1.196
	(1.434)	(1.423)	(1.425)	(1.408)	(1.412)
Relative size	1.325	1.318	1.294	1.292	1.300
	(0.753)	(0.744)	(0.693)	(0.688)	(0.704)
Constant	4.587**	4.467**	4.492**	4.540**	4.491**
	(2.281)	(2.258)	(2.263)	(2.279)	(2.269)
Observations	3,681	3,681	3,681	3,681	3,681
Pseudo R-squared	0.0182	0.0182	0.0182	0.0182	0.0182

Table 2.8: Hedge fund manager connections and equity portfolio weights

This table presents pooled OLS regression results of the equity portfolio weights on the connection variables. The sample comprises the quarterly equity holdings by the 430 sample hedge fund firms during the 2007-2015Q1 sample period. Stocks with missing or negative book equity and missing CRSP return data are excluded from the sample. The dependent variable is the fraction of dollars invested in a stock to the hedge fund firms total equity investment. The binary variables *Connection1* through *Connection4* are the key independent variables that capture the strength of connections between the sample hedge fund managers and the directors at their portfolio firms. *Connection1* takes the value 1 if any of the board members in a portfolio firm attended the same universities as the fund managers at a hedge fund firm that owns the stock and takes 0 otherwise. *Connected2* takes the value 1 if any of the board members in a portfolio firm attended the same universities and received the same degrees as the fund managers at a hedge fund firm that owns the stock and takes 0 otherwise. *Connection2B* takes the value 1 if any of the board members in a portfolio firm attended the same universities and finished their degrees in the same years as the fund managers at a hedge fund firm that owns the stock and takes 0 otherwise. *Connection3* takes the value 1 if any of the board members in a portfolio firm attended the same universities and received the same degrees in the same years as the fund managers at a hedge fund firm that owns the stock and takes 0 otherwise. *Connection4* takes the value 1 if any of the board members in a portfolio firm attended the same universities and received the same degrees in the same years as the top officials at a hedge fund firm that owns the stock and takes 0 otherwise. A degree year must precede the year of observation by at least two years to be considered as a connection. *Same state* is a binary variable that takes the value 1 if the headquarters of a hedge fund firm and of a portfolio firm is located in the same state and takes 0 otherwise. *Size*, *B/M*, and *Momentum* are percentiles of market value of equity, book-to-market value of equity, and the past 1 year stock performance. *Portfolio size* is the log-transformed total dollar value of the sample hedge fund firms total equity investment. All models include quarter fixed effects. Panel A shows the baseline estimation results. Panel B shows the results on

Connection3 and *Connection4* with additional fixed effects on the individual portfolio firms, on portfolio firms industries, and on the hedge fund firms. Panel C reports the regression results on *Connection3* and *Connection4* with varying distance between the degree years. Panel D reports the regression results on *Connection3* and *Connection4* with the senior directors and junior directors separately. We define senior directors as directors who assume any of the following positions: the chairperson of a board, the chief executive officer or the chief financial officer of a firm, a member of the executive committee, or a chairperson of any subcommittees of a board. Junior directors are defined as directors who assume none of these positions. Student t-statistics from standard errors clustered by quarter is reported in the parentheses. ***, **, * indicate significance at 1%, 5% and 10% levels, respectively.

Table 2.8: Hedge fund manager connections and equity portfolio weights

Panel A: Regression of Portfolio weights (in basis points)

Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Constant	645.1***	651.8***	643.8***	643.7***	643.5***	645.1***
	(79.54)	(79.66)	(80.31)	(80.33)	(80.13)	(79.48)
Connection1	2.619***					2.388**
	(2.968)					(2.723)
Connection2A		11.65***				
		(11.22)				
Connection2B			7.987***			
			(3.928)			
Connection3				16.91***		
				(7.566)		
Connection4					40.10***	38.95***
					(10.66)	(10.86)
Same state	14.46***	13.27***	14.78***	14.80***	14.78***	14.41***
	(17.62)	(15.85)	(16.60)	(16.53)	(16.43)	(17.49)
Size	0.738***	0.714***	0.745***	0.745***	0.745***	0.737***
	(28.02)	(27.39)	(28.60)	(28.66)	(28.63)	(27.98)
B/M	-0.0340**	-0.0346**	-0.0338**	-0.0338**	-0.0341**	-0.0344**
	(-2.334)	(-2.384)	(-2.315)	(-2.303)	(-2.321)	(-2.353)
Momentum	0.156***	0.159***	0.156***	0.156***	0.156***	0.156***
	(9.263)	(9.461)	(9.187)	(9.186)	(9.179)	(9.261)
Portfolio size	-44.33***	-44.91***	-44.18***	-44.17***	-44.16***	-44.32***
	(-68.23)	(-68.29)	(-69.19)	(-69.22)	(-69.04)	(-68.16)
Quarter FE	YES	YES	YES	YES	YES	YES
Observations	728,705	728,705	728,705	728,705	728,705	728,705
R-squared	0.126	0.126	0.126	0.126	0.126	0.126

Panel B: Regression of Portfolio weights (in basis points) with various FE

Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Constant	655.4***	651.3***	653.9***	655.2***	651.2***	653.9***
	(71.04)	(78.84)	(28.60)	(70.87)	(78.67)	(28.59)
Connection3	18.06***	16.68***	3.374**			
	(7.387)	(7.600)	(2.476)			
Connection4				43.11***	40.08***	7.513**
				(11.44)	(11.03)	(2.716)
Same state	13.17***	13.66***	3.929***	13.16***	13.65***	3.928***
	(22.04)	(18.36)	(5.262)	(21.89)	(18.23)	(5.250)
Size	0.475***	0.757***	0.777***	0.476***	0.757***	0.777***
	(6.913)	(32.65)	(67.05)	(6.934)	(32.62)	(67.00)
B/M	-0.128***	-0.0234	0.0116	-0.128***	-0.0237	0.0116
	(-6.382)	(-1.490)	(1.197)	(-6.406)	(-1.510)	(1.193)
Momentum	0.0788***	0.162***	0.161***	0.0787***	0.162***	0.161***
	(5.573)	(10.14)	(12.22)	(5.561)	(10.13)	(12.22)
Portfolio size	-42.42***	-43.97***	-44.42***	-42.41***	-43.96***	-44.42***
	(-71.64)	(-68.60)	(-26.04)	(-71.45)	(-68.43)	(-26.04)
Quarter FE	YES	YES	YES	YES	YES	YES
CUSIP FE	YES	NO	NO	YES	NO	NO
FF48 FE	NO	YES	NO	NO	YES	NO
HF Company FE	NO	NO	YES	NO	NO	YES
Observations	728,705	728,705	728,705	728,705	728,705	728,705
R-squared	0.193	0.129	0.464	0.193	0.129	0.464

Panel C: Varying graduation year threshold

Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Constant	643.7***	643.9***	644.1***	643.5***	643.7***	643.9***
	(80.33)	(80.50)	(80.68)	(80.13)	(80.24)	(80.31)
Connection3	16.91***					
	(7.566)					
Connection3 with		8.432***				
± 1 graduation year		(4.391)				
Connection3 with			8.000***			
± 2 graduation years			(4.486)			
Connection4				40.10***		
				(10.66)		
Connection4 with					26.30***	
± 1 graduation year					(7.071)	
Connection4 with						26.49***
± 2 graduation years						(6.671)
Quarter FE	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES
Observations	728,705	728,705	728,705	728,705	728,705	728,705
R-squared	0.126	0.126	0.126	0.126	0.126	0.126

Panel D: Varying firm executives seniority

Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Constant	643.7***	643.5***	643.7***	643.5***	643.4***	643.5***
	(80.27)	(80.19)	(80.33)	(80.15)	(80.12)	(80.14)
Connection3 with	21.51***		21.40***			
senior director	(10.07)		(9.995)			
Connection3 with		9.973**	9.685**			
junior director		(2.398)	(2.308)			
Connection4 with				48.68***		48.49***
senior director				(12.68)		(12.63)
Connection4 with					29.44***	29.07***
junior director					(4.705)	(4.596)
Quarter FE	YES	YES	YES	YES	YES	YES
Other controls	YES	YES	YES	YES	YES	YES
Observations	728,705	728,705	728,705	728,705	728,705	728,705
R-squared	0.126	0.126	0.126	0.126	0.126	0.126

Table 2.9: Hedge fund manager connections and equity returns

This table presents calendar time portfolio returns for the 2007-2015Q1 sample period. In Panel A, we assign each stock held by each sample hedge fund firm to either Connected or Unconnected portfolios at the beginning of each quarter. We then calculate monthly weighted average return of each portfolio assuming that the hedge fund firm maintains the same portfolio weights until the weights are updated at the end of each quarter, and compute the monthly weighted average return of each portfolio across all the sample hedge fund firms using the total dollar value of each hedge funds equity investment to the total equity investment of all the hedge fund firms as weights. Panel A display the intercepts (alpha) from the regressions of the monthly excess portfolio returns on the Fama and French (1993) three factors and the Carhart (1997) momentum factor; the monthly portfolio returns computed using the raw stock returns, averaged over the sample period; and the monthly portfolio returns computed using the Daniel, Grinblatt, Titman, and Wermers (1997) benchmark-adjusted stock returns, averaged over the sample period. In Panel B, we assign each CRSP stock whose directors have connections with fund managers in a sample hedge fund firm to either Connected & Held or Connected & Not Held portfolios at the beginning of each quarter. Next, we calculate monthly weighted average return of each portfolio assuming that the hedge fund firm maintains the portfolio weights until the end of each quarter. For the Connected & Not Held portfolio firms, we use the market value of equity of each stock to generate the portfolio weights. In Panel C, we divide the connected portfolio further into two sub-portfolios: Connected to senior directors and Connected to junior directors. Panel D reports the results from the same tests an in Panel A but only on the stocks without same state connections with the hedge fund firms.

Table 2.9: Hedge fund manager connections and option performance

Panel A: Held & Connected versus Held & Unconnected

	4 Factor alpha			Raw returns			Benchmark-adj. returns			Sharpe Ratio	
Connection	**Conn.**	**Uncon.**	**Long-short**	**Conn.**	**Uncon.**	**Long-short**	**Conn.**	**Uncon.**	**Long-short**	**Conn.**	**Uncon.**
Connection1	-0.0007	0.0003	-0.0010	0.0076	0.0091	-0.0015	0.0004	0.0019**	-0.0015**	0.1329	0.1561
	(0.8626)	(0.2503)	(1.1861)	(1.4118)	(1.6371)	(-1.6038)	(0.7489)	(2.1574)	(-2.0739)		
Connection2A	-0.0004	-0.0001	-0.0003	0.0078	0.0085	-0.0007	0.0005	0.0013	-0.0008	0.1358	0.1470
	(0.5703)	(0.1103)	(0.4023)	(1.4396)	(1.5483)	(-0.8288)	(0.9168)	(1.6386)	(-1.1093)		
Connection2B	0.0026**	-0.0004	0.0030**	0.0105*	0.0081	0.0025*	0.0030***	0.0009	0.0021*	0.1873	0.1400
	(2.2597)	(0.4265)	(2.1955)	(1.9491)	(1.4810)	(1.8346)	(2.6664)	(1.5807)	(1.9333)		
Connection3	0.0049***	-0.0003	0.0053***	0.0128**	0.0081	0.0047**	0.0046***	0.0010	0.0036**	0.2093	0.1411
	(3.5288)	(0.3743)	(3.3912)	(2.1564)	(1.4915)	(2.5553)	(3.0225)	(1.6397)	(2.3039)		
Connection4	0.0077***	-0.0003	0.0080***	0.0148**	0.0081	0.0067***	0.0058***	0.0010*	0.0048***	0.2499	0.1414
	(4.3389)	(0.3562)	(3.8731)	(2.5597)	(1.4947)	(2.8112)	(3.3260)	(1.6697)	(2.6519)		

Panel B: Held & Connected versus Not Held & Connected

	4 Factor alpha			Raw returns			Benchmark-adj. returns			Sharpe Ratio	
Connection	**Conn.**	**Uncon.**	**Long-short**	**Conn.**	**Uncon.**	**Long-short**	**Conn.**	**Uncon.**	**Long-short**	**Conn.**	**Uncon.**
Connection1	-0.0007	0.0001	-0.0007	0.0076	0.0077	-0.0000	0.0004	0.0001	0.0003	0.1329	0.1340
	(0.8626)	(0.1542)	(0.9094)	(0.1612)	(0.1580)	(0.9807)	(0.7489)	(0.1507)	(0.3805)		
Connection2A	-0.0004	0.0000	-0.0004	0.0078	0.0076	0.0002	0.0005	0.0000	0.0005	0.1358	0.1324
	(0.5703)	(0.0115)	(0.5372)	(0.1532)	(0.1627)	(0.8567)	(0.9168)	(0.0478)	(0.5943)		
Connection2B	0.0026**	0.0003	0.0023*	0.0105*	0.0083	0.0022*	0.0030***	0.0009	0.0021*	0.1873	0.1384
	(2.2597)	(0.3179)	(1.9562)	(0.0542)	(0.1471)	(0.0950)	(2.6664)	(0.8833)	(1.7265)		
Connection3	0.0049***	0.0002	0.0047***	0.0128**	0.0082	0.0046***	0.0046***	0.0007	0.0040**	0.2093	0.1299
	(3.5288)	(0.2077)	(2.9974)	(0.0335)	(0.1727)	(0.0041)	(3.0225)	(0.5585)	(2.5507)		
Connection4	0.0077***	0.0005	0.0072***	0.0148**	0.0082	0.0065***	0.0058***	0.0009	0.0048**	0.2499	0.1275
	(4.3389)	(0.3615)	(3.4378)	(0.0120)	(0.1813)	(0.0025)	(3.3260)	(0.6174)	(2.4384)		

Panel C: Connections with senior directors versus connections with junior directors

C1: Senior director connection

Connection	**4 Factor alpha**	**Raw return**	**Adj. return**
Connection1	-0.0005	0.0078	0.0006
	(0.6218)	(1.4285)	(0.9619)
Connection2A	-0.0006	0.0077	0.0002
	(0.7372)	(1.3879)	(0.3415)
Connection2B	0.0022	0.0104*	0.0038**
	(1.4322)	(1.8755)	(2.3775)
Connection3	0.0048**	0.0128**	0.0052***
	(2.5589)	(2.1209)	(2.7683)
Connection4	0.0071***	0.0138**	0.0056**
	(2.6992)	(2.3736)	(2.1798)

C2: Junior director connection

Connected	**4 Factor alpha**	**Raw return**	**Adj. return**
Connection1	-0.0006	0.0075	0.0004
	(1.0107)	(1.4195)	(0.6746)
Connection2A	-0.0003	0.0079	0.0007
	(0.5020)	(1.4633)	(1.2429)
Connection2B	0.0028*	0.0105*	0.0027*
	(1.8324)	(1.9266)	(1.8531)
Connection3	0.0034	0.0107*	0.0022
	(1.5668)	(1.8278)	(1.0938)
Connection4	0.0053*	0.0132*	0.0037
	(1.9164)	(1.9378)	(1.2417)

Panel D: Held & Connected with no same state observations

Connection	**4 Factor alpha**	**Raw return**	**Adj. return**
Connection1	-0.0005	0.0078	0.0006
	(0.6646)	(1.4422)	(0.9499)
Connection2A	-0.0003	0.0079	0.0007
	(0.3875)	(1.4689)	(1.1297)
Connection2B	0.0023*	0.0104*	0.0031**
	(1.9105)	(1.9163)	(2.5870)
Connection3	0.0049***	0.0130**	0.0053***
	(3.0702)	(2.1799)	(3.1406)
Connection4	0.0065***	0.0141**	0.0057***
	(3.4865)	(2.3709)	(3.1621)

BIBLIOGRAPHY

[1] Ajay Agrawal, Devesh Kapur, and John McHale. How do spatial and social proximity influence knowledge flows? Evidence from patent data. Journal of Urban Economics, 64(2), 2008.

[2] Kenneth R. Ahern. Information networks: Evidence from illegal insider trading tips. Journal of Financial Economics (forthcoming).

[3] George O. Aragon and J. Spencer Martin. A unique view of hedge fund derivatives usage: Safeguard or speculation? Journal of Financial Economics, 105.

[4] Paul Asquith, Michael B. Mikhail, and Andrea S. Au. Information content of equity analyst reports. Journal of Financial Economics, 75(2):245–282, 2005.

[5] Patrick Augustin, Menachem Brenner, and Marti G. Subrahmanyam. Informed options trading prior to M&A announcements: Insider trading. Working paper, 2015.

[6] Jan Bena and Kai Li. Corporate innovations and mergers and acquisitions. Journal of Finance, 69(5):1923–1960, 2014.

[7] Ola Bengtsson and David H. Hsu. How do venture capital partners match with startup founders? Working paper, 2010.

[8] Ola Bengtsson and David H. Hsu. Ethnic matching in the U.S. venture capital market. Journal of Business Venturing, 30(2):338–354, 2015.

[9] Allen Berger, Nathan H. Miller, Mitchell Petersen, Raghuram Rajan, and Jeremy Stein. Does function follow organizational form? Evidence from the lending practices of large and small banks.

[10] Cameron Blevins and Lincoln Mullen. Jane, John ... Leslie? A historical method for algorithmic gender prediction. Digital Humanities Quarterly, 9(3), 2015.

[11] Jeffrey R. Brown, Zoran Ivković, Paul A. Smith, and Scott Weisbenner. Neighbors matter: Causal community effects and stock market participation. Journal of Finance, 63(2).

[12] Tianqi Chen. Introduction to boosted trees. University of Washington teaching note, 2014.

[13] John Cochrane. The risk and return of venture capital. Journal of Financial Economics, 75(1):3–52, 2006.

[14] Lauren Cohen, Andrea Frazzini, and Christopher J. Malloy. The small world of investing: Board connections and mutual fund returns. Journal of Political Economy, 116(5):951–979, 2008.

[15] Lauren Cohen, Andrea Frazzini, and Christopher J. Malloy. Sell-side school ties. Journal of Finance, 65(4):1409–1437, 2010.

[16] Lily Fang and Joel Peress. Media coverage and the cross-section of stock returns. Journal of Finance, 64(5):2023–2052, 2009.

[17] Raymond J. Fisman. Ethnic ties and the provision of credit: Relationship-level evidence from african firms. Advances in Economic Analysis & Policy, 3(1):1–21, 2003.

[18] Vance H. Fried and Robert D. Hisrich. Towards a model of venture capital investment decision making. Financial Management, 23(3):28–37, 1994.

[19] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5):1189–1232, 2001.

[20] Jerome H. Friedman. Stochastic gradient boosting. Computational Statistics & Data Analysis, 38(4):367–378, 2002.

[21] Edward L. Glaeser, David Laibson, Jose Scheinkman, and Christine L. Soutter. Measuring trust. The Quarterly Journal of Economics, 115(3):811–846, 2000.

[22] Paul A. Gompers and Josh Lerner. The venture capital revolution. Journal of Economic Perspectives, 15(2):145–168, 2001.

[23] Paul A. Gompers, Vladimir Mukharlyamov, Emily Weisburst, and Yuhai Xuan. Gender effects in venture capital. Working paper, 2014.

[24] Paul A. Gompers, Vladimir Mukharlyamov, and Yuhai Xuan. The cost of friendship. Journal of Financial Economics, 119(3):626–644, 2016.

[25] John M. Griffin and Jin Xu. How smart are the smart guys? A unique view from hedge fund stock holdings. Review of Financial Studies, 22(7):2531–2570, 2009.

[26] Jarrad Harford, Robert Schonlau, and Jared Stanfield. Trade relationships, indirect economic links, and mergers. Working paper, 2016.

[27] Deepak Hegde and Justin Tumlinson. Does social proximity enhance business partnerships? Theory and evidence from ethnicity's role in U.S. venture capital. Management Science, 60(9):2355–2380, 2014.

[28] Thomas Hellman and Manju Puri. The interaction between product market and financing strategy: The role of venture capital. Review of Financial Studies, 13(4):959–984, 2000.

[29] Thomas Hellman and Manju Puri. Venture capital and the professionalization of start-up firms: Empirical evidence. Journal of Finance, 57(1):169–197, 2002.

[30] Yael Hochberg, Alexander Ljungqvist, and Yang Lu. Whom you know matters: Venture capital networks and investment performance. Journal of Finance, 62(1):266–301, 2007.

[31] Geert Hofstede. Culture's Consequences: International Differences in Work-related Values. Sage, Beverly Hills, CA, 1st edition, 1980.

[32] Geert Hofstede. Cultures and Organizations: Software of the Mind. McGraw-HIll, USA, 1st edition, 1997.

[33] Geert Hofstede. Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations. Sage, Thousand Oaks, CA, 2nd edition, 2001.

[34] Geert Hofstede, Gert J. Hofstede, and Michael Minkov. Cultures and Organizations: Software of the Mind. McGraw-Hill, USA, 3rd edition, 2010.

[35] Harrison Hong, Jeffrey D. Kubik, and Jeremy C. Stein. Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers. Journal of Finance, 60(6):2801–2824, 2005.

[36] David H. Hsu. Experienced entrepreneurial founders, organizational capital, and venture capital funding. Research Policy, 36(5):722–741, 2007.

[37] Hans K. Hvide and Per Östberg. Stock investments at work. Journal of Financial Economics, 117(3):628–652, 2015.

[38] Joy Ishii and Yuhai Xuan. Acquirer-target social ties and merger outcomes. Journal of Financial Economics, 112(3):344–363, 2014.

[39] G. Andrew Karolyi. The gravity of culture for finance. Journal of Corporate Finance, 41:610–625, 2016.

[40] William R. Kerr. Ethnic scientific communities and international technology diffusion. Review of Economics and Statistics, 90(3):518–537, 2008.

[41] William R. Kerr, Josh Lerner, and Antoinette Schoar. The consequences of entrepreneurial finance: Evidence from angel financings. Review of Financial Studies, 27(1):20–55, 2014.

[42] April Klein, Anthony Saunders, and Yu Ting Forester Wong. Do hedge funds trade on private information? Evidence from upcoming changes in analysts stock recommendations. Working paper, 2014.

[43] Bruce Kogut and Harbir Singh. The effect of national culture on the choice of entry mode. Journal of International Business Studies, 19(3):411–432, 1988.

[44] Josh Lerner. The syndication of venture capital investments. Financial Management, 23(3):16–27, 1994.

[45] Josh Lerner. Venture capitalists and the oversight of private firms. Journal of Finance, 50(1):301–318, 1995.

[46] Nadia Massoud, Debarshi Nandy, Anthony Saunders, and Keke Song. Do hedge funds trade on private information? Evidence from syndicated lending and short-selling. Journal of Financial Economics, 99(3):477–499, 2011.

[47] Pablo Mateos. A review of name-based ethnicity classification methods and their potential in population studies. Population Space and Place, 13(4):243–263, 2007.

[48] Miller McPherson, Lynn Smith-Lovin, and James M. Cook. Birds of a feather: Homophily in social networks. Annual Review of Sociology, 27:415–444, 2001.

[49] Bang Dang Nguyen. Does the rolodex matter? Corporate elite's small world and the effectiveness of boards of directors. Management Science, 58(2):236–252, 2012.

[50] Duc Duy Nguyen, Jens Hagendorff, and Arman Eshraghi. Does CEO cultural heritage affect performance under competitive pressure? Review of Financial Studies (forthcoming).

[51] Han N. Ozsoylev, Johan Walden, M. Deniz Yavuz, and Recep Bildik. Investor networks in the stock market. Review of Financial Studies, 27(5):1323–1366, 2014.

[52] Yihui Pan, Stephan Siegel, and Tracy Yue Wang. Corporate risk culture. Journal of Financial and Quantitative Analysis (forthcoming), 2016.

[53] Yihui Pan, Stephan Siegel, and Tracy Yue Wang. The cultural origin of preferences: CEO cultural heritage and corporate acquisitions. Working paper, 2016.

[54] Sahil Raina. VC financing and the entrepreneurship gender gap. Working paper, 2017.

[55] Martin Ruef, Howard E. Aldrich, and Nancy M. Carter. The structure of founding teams: Homophily, strong ties, and isolation among U.S. entrepreneurs. American Sociological Review, 68(2):195–222, 2003.

[56] Social Security Administration. National baby names. `https://www.ssa.gov/OACT/babynames/limits.html`, 2016. Online; accessed 10 September 2016.

[57] Morten Sørenson. How smart is smart money? A two-sided matching model of venture capital. Journal of Finance, 62(6):2725–2762, 2007.

[58] Jeremy C. Stein. Conversations among competitors. American Economic Review, 98(5):2150–2162, 2008.

[59] TechCrunch. Crunchbase 2013 snapshot. `https://data.crunchbase.com/docs/2013-snapshot`, 2013. Online; accessed 18 June 2016.

[60] D. Kenneth Tucker. The Cultural-Ethnic-Language Group technique as used in the Dictionary of American Family Names (DAFN).

[61] D. Kenneth Tucker. Distribution of forenames, surnames, and forename-surname pairs in the United States. Names, 49(2):69–96, 2001.

[62] U.S. Attorney, Southern District of New York. Former Intel executive pleads guilty in Manhattan federal court to insider trading with Galleon Hedge Fund manager. `https://www.justice.gov/archive/usao/nys/pressreleases/February10/goelrajivpleapr.pdf`, 2010. Online; accessed 4 February 2017.

[63] Eric Van den Steen. On the origin of shared beliefs (and corporate culture). RAND Journal of Economics, 41(4):617–648, 2010a.

[64] Eric Van den Steen. Culture clash: The costs and benefits of homogeneity. Management Science, 56(10):1718–1738, 2010b.

[65] Xinxin Wang. Catering innovation: Entrepreneurship and the acquisition market. Working paper, 2016.

[66] World Values Survey Association. World Values Survey 1981-2014 longitudinal aggregate. `http://www.worldvaluessurvey.org`, 2015. Online; accessed 13 March 2017.

[67] Stefan Wuyts and Inge Geyskens. The formation of buyer-supplier relationships: Detailed contract drafting and close partner selection. Journal of Marketing, 69(4):103–117, 2005.

[68] Guang Xiang, Zeyu Zheng, Miaomiao Wen, Jason Hong, Carolyn Rose, and Chao Liu. A supervised approach to predict company acquisition with factual and topic features using profiles and news articles on TechCrunch. Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, 2012.

Appendix A

GRADIENT BOOSTED TREES

In the Gradient Boosted Trees method (Friedman, 2001; 2002), the target function is modeled as an ensemble of Classification and Regression Trees (CART), by which the outcome variable is related to a set of input variables through a set of classification questions on the input variables.

An example of how a CART works follows: consider a database that contains variables covering the age, gender, occupation, and other characteristics of a family of three with a mother and her daughter and son. This database is used to answer a question of interest, which in this case could be whether the person enjoys computer games. At its first branch, the CART could split the sample on the basis of age. If those above the age of thirty (in this case, the mother) do not enjoy computer games, they would form a terminal node or leaf with a negative prediction score. At the next branch, the CART would therefore only consider the children. The tree could then next split the sample on the basis of whether the person is male, and assign the son to a leaf with a positive prediction score if he does enjoy computer games, and the daughter to another leaf with another leaf with a smaller positive prediction score if she also enjoys computer games, but not as much as her brother.

However, a single CART is generally insufficient on its own, as there may be other specific configurations or combinations of splits or branches that were not considered in a single specific CART but could still provide additional information for classification. For example, a second CART could instead split the sample based on whether the person uses the computer daily, and the number of hours that the person uses the computer each day. In order to consider these other possibilities, the tree ensemble model instead uses multiple CARTs together in an ensemble to ensure that many possible configurations of the input

variables or splits are considered. The prediction scores of the multiple CARTs are then summed up. The machine then learns the best CARTs to include by defining an objective function based on the predictive power and the simplicity, as opposed to complexity, of the CARTs it uses in the learning process.