NBER WORKING PAPER SERIES

HOMOPHILY IN ENTREPRENEURIAL TEAM FORMATION

Paul A. Gompers
Kevin Huang
Sophie Q. Wang

Working Paper 23459
http://www.nber.org/papers/w23459

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2017

Support for this research was provided by the Division of Research at the Harvard Business School. Lauren Cohen provided helpful comments and suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Homophily in Entrepreneurial Team Formation
Paul A. Gompers, Kevin Huang, and Sophie Q. Wang
NBER Working Paper No. 23459
May 2017
JEL No. J15,J16

ABSTRACT

We study the role of homophily in group formation. Using a unique dataset of MBA students, we observe homophily in ethnicity and gender increases the probability of forming teams by 25%. Homophily in education and past working experience increases the probability of forming teams by 17% and 11 % respectively. Homophily in education and working experience is stronger among males than females. Further, we examine the causal impact of homophily on team performance. Homophily in ethnicity increases team performance by lifting teams in bottom quantiles to median performance quantiles, but it does not increase the chance of being top performers. Our findings have implications for understanding the lack of diversity in entrepreneurship and venture capital industry.

Paul A. Gompers
Harvard Business School
Baker Library 263
Soldiers Field
Boston, MA 02163
and NBER
pgompers@hbs.edu

Kevin Huang
Baker Library
Harbard Business School
Boston, MA 02163
kahuang@hbs.edu

Sophie Q. Wang
Department of Economics
Harvard University
Littauer Center
1805 Cambridge Street
Cambridge, MA 02138
USA
sophieqzwang@gmail.com

Contents

1. Introduction 3

2. Setting 6

3. Data 8

4. Empirical Results on Matching 9

4.1 Ethnicity Homophily 11

4.2 Gender Homophily 13

4.4 Education Homophily 14

4.5 Past Industry Experience Homophily 15

5. Homophily and Performance 16

6. Conclusion and Discussion 22

1. Introduction

Literature in sociology has documented the existence of homophily, the desire to associate with those similar to you, in various social networks, from the strongest social ties such as in marriage (Kalmijn 1998, Fiore and Donath 2005), close friendships (Marsden, 1987, 1988, Currarini, Jackson and Pin 2009), to professional networks (Gompers, Muhkarlyamov, and Xuan 2016; Ruef, Aldrich and Carter 2003; Sorenson and Stuart 2001) and acquiescence (Hampton and Wellman 2000). Homophily can arise from the similarities in demographic characteristics, such as ethnicity, country of origin, age, and gender. It can also be based on acquired characteristics, such as education, occupation and religion (Lazarsfeld and Merton 1954, Mcpherson et al. 2001). Most past research has focused on homophily in ethnicity and gender. Relatively few studies[1] have examined homophily in educational and professional backgrounds due to limitation in data. Verbrugge (1977) and Louch (2000) explore homophily in both demographic and socioeconomic characteristics. While they confirm the existence of homophily along both dimensions, they do not distinguish the relative strength of homophily that each dimension engenders.

In this paper, we contribute four major findings to the literature on homophily and its effects on performance. First, we estimate the relative economics magnitudes of homophily in ethnicity, gender, education and working experience. The first central question we address is: what are the strongest homophilic forces in forming social networks? Using a novel dataset of HBS MBA students, we find ethnicity and gender are the two strongest homophilic forces in social networks. Individuals are 25% more likely to form groups with people of the same ethnicity or gender relative to randomly matching within a set of students who choose the groups that they work with on real microbusinesses. Homophily in education and working experience is weaker than demographic

[1] Among studies that do include education homophily, most of them use "Education year" instead of past educational institution as a dimension of homophily.

homophily, but they are still economically significant. School ties and shared working experience increase the probability of forming social networks by 17% and 11% respectively. Further, we find homophily in education and working experience is stronger among males than females.

Our second contribution is to examine the relationship between group size and strength of homophily. Currarini, Jackson, and Pin (2009) study the friendship patterns in American high schools. They find the existence of homophily among all ethnic groups, and it is the strongest among middle-sized groups. They present an economic model in which homophily is generated by bias in preference and bias in the meeting process. In a follow-up paper by the same authors (Currarini, Jackson, and Pin, 2010), they empirically estimate the magnitude of each source of homophily for different ethnic groups. They find that Asians and African American exhibit the strongest biases in the meeting process. A slightly different model by Currarini and Redondo (2016) predicts that homophily only exists in relatively large groups, because the cost of inbreeding will be too high for very small groups.

Consistent with Currarini and Redondo (2016), we find homophily is weaker among Hispanic American and African American students, two relatively underrepresented groups compared to White Americans and Asian Americans in our sample of Harvard MBA students. This could be due to the high search cost within small groups, or as a result of the strategic decisions to team up with White and Asian Americans as a means to compensate for being underrepresented minorities. We also find homophily is strongest among international students, students who graduated from non-ivy league schools, and students who worked in non-major industries[2]. This is contradictory to Currarini and Redondo's model. The group sizes of these students are small, yet the

[2] Non major industries refer to non-finance, non-consulting and non-technology industries. Most of the students are from finance, consulting or technology industries.

homophily is the strongest. Future theoretical work should explore the interaction between group size and different types of homophilic ties with heterogeneous strengths.

Third, our unique dataset allows us to explore the dynamics of entrepreneurial team formation. Gompers and Wang (2017) document the homogeneity in gender and ethnicity in US start-up teams. Female and non-Asian minorities have been underrepresented in the innovation sector for the past 20 years and the progress to achieve diversity has been slow. One possible explanation is the biases of hiring people with similar background. Despite large volumes of research on homophily in various settings, only a few studies have explored the effect of homophily in entrepreneurship. Ruef, Ruef, and Carter (2003) survey 830 entrepreneurs on their founding team composition. They find that the probability of a team with the same gender or with the same ethnicity is higher than a random matching process would predict. In our setting, we observe MBA students tend to form entrepreneurial teams with people who have similar social and demographic backgrounds. Given a significant portion of students will be working at start-ups and venture capital firms after graduation, our results have implications for understanding start-up team diversity, recruitment process in start-ups and venture capital firms, and deal selection in venture capital.

Finally, our paper contributes to the understanding of the causal relationship between team diversity and performance. Theoretical work on diversity focuses on the trade-off between the information gains and the communication costs. Heterogeneous teams benefit from more diverse pools of skill and knowledge, but at the same time, differences in ethnicity, culture, and mother language hinder efficient communication among team members, thus potentially lowering productivity. (Alesina and La Ferrara 2003, Lazear 1999). Knippenberg and Schipper (2007) review empirical literature on team diversity and performance from 1997 to 2005, and they conclude that the empirical results on diversity are "highly inconsistent" because of the endogenous process of group formation.

Recent studies use field experiment to alleviate the endogeneity concern. Hoogendoorn and Praag (2012) find the benefit of information sharing is greater than communication cost in more ethnically diverse teams. Marx et al. (2015) find horizontal diversity (i.e., at the same level of authority) in ethnicity decreases team efficiency, because people in heterogeneous teams are more likely to complain about their teammates. Vertical heterogeneity (i.e., at different levels of authority), on the contrary, increases team performance, as workers tend to exert more effort when the manager is from a different ethnic background. Gompers and Wang (2017) find parenting daughter increases venture partners' tendency to hire female investment partners. Using the number of daughters by senior venture partners as the instrument for venture capital firm gender diversity, the authors find gender diversity improves venture capital firm's investment performance. Our study provides a clean setting to test the causal relationship between diversity and team performance. By exploiting a quasi-experimental setting of team assignments in the class year of 2013, we find homogeneity in ethnicity increases team performance by lifting teams from the bottom quantiles to median performance quantile, potentially because it increases communication efficiency and lowers the probability of conflict within the team. However, homogeneity does not increase the chance of being top performers. We do not find homogeneity in gender, education, or past work experience is an important factor in determining team performance.

2. Setting

First year MBA students at the HBS from 2012 through 2016 were required to take a field course in the spring semester of their first year. Throughout the course, students were required to design and launch a real microbusiness. At the beginning of the semester, students formed teams of

5-7 people from the same section.[3] Two months into the semester, students presented their projects to faculty members. If the faculty members believed the proposed project was achievable, the team then proceeded to present their project to a panel of judges at the end of semester ("IPO day"). The panel of judges then ranked all the projects based on teams' performance and the quality of the idea during the "IPO day".

When the field course was first introduced to the students in the spring semester of 2012 for the MBA Class of 2013[4], the school assigned each student to the teams based on their background. One goal of the assignments was to make teams were somewhat diverse in terms of gender, ethnicity, education, and past working experience. After 2013, the school changed the team formation policy and started to have students choose teammates themselves. The school did not impose any restriction on how students formed their teams. Anecdotal evidence suggests that students frequently formed teams with friends who had similar demographic backgrounds. Figure 1 plots the probability of a student being matched to her classmate conditional on having the same ethnicity, gender, education, or industry backgrounds. The conditional probability of matching increases in all four dimensions when students are allowed to find teammates freely. This provides clean evidence on the existence of homophily during the process of team formation. In the next sections, we explore the relative strength of homophily based upon ethnicity, gender, education and past industry experience. From there, we explore the performance implications of diversity on performance.

Because teams were assigned by the MBA Administration for the Class of 2013, the diversity of teams is exogenous to each team member. As such, the causal implications of diversity for

---

[3] Harvard Business School students are assigned to one of ten sections in their first year and take all of their classes with the same roughly 90 students.

[4] 2013 refers to the class year of 2013, so do 2014, 2015 and 2016 later in the paper. Students take the field course at first year. Eg. Class year 2013 students take field course in 2012.

performance can be estimated for the Class of 2013. We also explore the performance impact of diversity for the Classes of 2014-2016, although endogeneity of team diversity makes interpretation of the performance results difficult.

3. Data

Our data were provided by the HBS MBA Program. In the data, we observe the gender, ethnicity, home country, undergraduate institution, past employer, and the industry of each MBA student from class year 2013 to 2016 but were not provided with students' actual names. Table I reports the summary statistics on 3,684 MBA students in our sample. Females make up 40% of total student population. Approximately 40% of the students are white Americans, 12% are Asian Americans, 5% are African Americans, 4% are Hispanic Americans, and 35% are international students. India, Canada, and China represent the top three origin countries for international students[5]. In terms of past work experience, roughly half of the students are from the finance or consulting industries, and not surprisingly, the big three consulting firms (McKinsey, Bain and BCG) and the two largest investment banks (Goldman and Morgan Stanley) are the top 5 suppliers of Harvard MBA students (Table II). Approximately 11% of students had experience in the technology industry, and this number increased by more than 50% from 2013 to 2016. 27% of the MBA students graduated from Ivy League schools. Harvard, Stanford, and University of Pennsylvania are the top 3 undergraduate institutes (Table II).

We also observe the team selection of each student. From 2013 to 2015, there are 150 teams in each class year and the average team size is 6. In 2016, the average team size was changed to 5 and there were 180 teams. To examine the effects of homophily on team formation, we construct

[5] Online Appendix Table 1.

student-student pairs by matching each student to every other student within the same section and year. This process creates 335,686 potential pairs. We then create a dependent variable "real_match" which equals to 1 if the two students are members of the same team and 0 otherwise. The independent variable "ethnic (gender, education, industry) tie" equals to 1 if two students belong to the same ethnic (gender, education, industry) group. Our data construction method is similar to Louch (2000). To illustrate, consider the following example: James Brown is a Section A student in 2013, and he has 5 teammates. We match Mr. Brown to all his section mates (89 of them) by creating 89 student-student pairs. Intuitively, each pair is a potential teammate with whom Mr. Brown could be paired. If the match happened randomly, Mr. Brown would pair with an arbitrary teammate with a probability of 5.6%[6]. Variable "real_match" equals 1 for the 5 pairs for which Mr. Brown is matched to his real teammates. To measure the effect of homophily on matching, we compare the probability of matching conditional for a pair having the same ethnicity (Gender, Education and Industry) to the probability of matching for a pair with different ethnicities (Gender, Education and Industry). Our baseline results are estimated using the following regression models:

$$Real\ Match_i = b_{11} * \ Ethnicity\ Tie_i + b_{12} * Team\ Size + Year\ FE + e_i$$

$$Real\ Match_i = b_{21} * \ Gender\ Tie_i + b_{22} * Team\ Size + Year\ FE + e_i$$

$$Real\ Match_i = b_{31} * \ Education\ Tie_i + b_{32} * Team\ Size + Year\ FE + e_i$$

$$Real\ Match_i = b_{41} * \ Industry\ Tie_i + b_{42} * Team\ Size + Year\ FE + e_i$$

4. Empirical Results on Matching

In this section, we examine the relative strength of homophily for ethnicity, gender, educational background and past work experience. While homophily is an economically significant

[6] 5/89=5.6%

force across all four dimensions, it is strongest in demographic characteristics, namely ethnicity and gender. Table III Panel A presents the regression results for matching from 2014 to 2016, the years for which students were allowed to choose their own teams. "Ethnicity Tie" increases the probability of matching by 1.38%. Given the base rate of matching is 5.6%, this represents a 25% increase from the baseline probability of randomly matching with a student of the same ethnicity. Similarly, we find common gender increases the probability of matching by 1.33%. The effect of homophily for education and past industry experience is smaller compared to homophily for gender and ethnicity. Attending the same undergraduate institution increases the probability of matching by 0.976%, a 17% increase from the baseline, and having the same industry experience increases the matching rate by 0.637%, an 11% increase from the baseline. Both these results are significant and economically meaningful. Panel B reports the regression result using 2013 subsample. We do not expect there to be any homophilic effects because teams were assigned in an attempt to promote diversity. The coefficients on "ethnicity tie", "school tie" and "industry tie" are negative and close to zero. The coefficient on "gender tie" is -1.67% and statistically significant at 1% level. The matching rate is much lower among student pairs of the same gender compared to student pairs of different genders. This reflects HBS's assignment scheme, which appears to match males to females to balance the gender ratio within each team. Interestingly, other dimensions do not seem to have been important in assigning teams.

Our results are largely consistent with prior research on homophily. McPherson and Smith (2001) give a comprehensive review on homophily in social networks. It is well documented that homophily exists in both demographic characteristics and acquired characteristics. Verbrugge (1977) provides some early evidence that homophily bias is stronger in demographic characteristics. To our best knowledge, our study is the first attempt to estimate and compare the relative strength of homogeneous ties in group formation.

4.1 Ethnicity Homophily

While common ethnicity is one of the strongest homophilic forces, its strength varies across different ethnic groups. We attempt to look at how the relative size of the ethnic group influences how strong the attraction is. We find homophily is strongest among international students, followed by Asian Americans and White Americans. It is relatively weaker among African Americans and non-existent among Hispanic Americans. As discussed below, this could be due to the high cost of searching among small groups or as a result of strategic decision making by underrepresented minorities.

In Table IV, the first two columns show that homophily increases the probability of matching by 1.08% and 1.16% among White American and Asian American MBA students repectively. Given the base rate of matching is 5.6%, this represents a 20% increase in the matching rate. The coefficient for African American students is 0.96%, but it is not statistically significant. Homophily has no effect on Hispanic American MBA students. Breaking down the matching rate by year (Online Appendix Table 2), we observe large variance among Hispanic American students. The matching rate was 11.29% among Hispanic Americans in 2014, and it is twice as large as the sample average (5.6%). However, the matching rate drops to 3.7% and 0% in 2015 and 2016. The large variance in matching rates may be due to the small population size of Hispanic American students in each class year. Given there are only 3.8 Hispanic American students in each section, they may not be able to find a teammate with same ethnic background easily. Similarly, the average number of African American students in each section is 5, and homophily among African American MBA students is relatively weak. This is consistent with Currarini and Redondo (2016)'s model, which predicts that matching is less likely to happen in small groups because the cost to do so is high. An alternative, and not exclusive, mechanism could be strategic decision making by underrepresented minorities. African American students and Hispanic American students may intentionally form teams with White American and Asian American students to compensate for the disadvantage of being underrepresented minorities.

Homophily is strongest among international students. An international MBA student is 3.77% more likely to find a teammate from the same region[7], three times greater than the effect among White and Asian Americans. A detailed breakdown of international students by region (Online Appendix Table 3) shows that homophily is strongest among students from East Asia, the Middle East, and Latin America. The coefficients on these three groups are around 6%, twice as large as the coefficients on Europeans and South Asian students. Note that there are only 3-4 students from East Asia, Middle East and Latin America per section, the strength of homophilic ties is considerably strong among these very small groups than it is for African American or Hispanic students who have a similar number of students per section. The coefficient for African students (Non-American) is negative and significant. This is because there is only approximately 1 African student in each section in each year.

Our matching results shed light on the interaction between the attraction of homophily and group size effect in matching. Group size can affect the matching process by increasing the cost of searching among minorities as in the case of African American and Hispanic American students. On the other hand, we also observe that homophily is strongest among international students, which is inconsistent with Currarini and Redondo (2016)'s model. Given the average group size of

[7] For other international students, we categorize their home countries by regions: Europe (7.7 students per section), South Asia (6.1 students per section), East Asia (4 students per section), Latin America (4 students per section), Middle East (3.3 students per section), Africa (1.6 students per section). Two exceptions are Canadians and Australians, we counted them also as white Americans (Online Appendix Table I).

internationals students is small[8], the cost of searching is very high, yet international students still tend to form groups with people from the same region regardless of the search cost.

4.2 Gender Homophily

Gender is another important homophilic factor in social network and gender homophily is stronger among female students than it is among male students. Table V shows that gender homophily increases the matching rate among females by 1.22%, and it is 68% higher than its effect on male students. Not surprisingly, the coefficient on gender homophily is negative and significant for both males and females in 2013, reflecting the group assignment scheme used by the school that was intended to increase gender diversity in teams.

Table VI breaks down ethnicity homophile by gender. The interaction between gender and ethnicity yields some interesting results. On average, males are more likely to form teams with people from the same ethnic background. The first and third column shows that "ethnicity tie" increases the probability of match by 1.54% among males and 1.14% among females. More specifically, white male students are 50% more likely to choose to form a team with another white male student than white female students are to form a team with another white female student. Among Hispanic American and international students, ethnicity homophily is also stronger among males than females. African American female students, on the contrary, have a higher probability of matching to another African American female student than are male students. The coefficient for African American female students is 2.45%, while the coefficient for African American male students is only 0.326%. Homophily almost does not exist among African American male students.

[8] On average, only 4-5 of international students are from the same region in each section

The interaction between gender and ethnic is less well understood (Block and Grund 2014, Wimmer and Lewis 2010), as previous studies often treat gender and ethnicity as separate categorizes. The above results suggest the lack of diversity in entrepreneurship is not a simple problem of one gender or one ethnicity. It is a more complex story about the interactions of gender and ethnicity. Policies that fail to consider this interaction effects may be effective in one part of the population but futile for the rest.

4.4 Education Homophily

Homophily also exists among people who share similar education background. Individuals are more likely to interact with people with same level of education (Verbrugge 1977; Louch 2000; Marsden 1988). People form long-term friendships with their classmates (Neckerman 1996). Equity analysts are more likely to build relationship and acquire superior information through school ties with the management (Cohen, Frazzini and Malloy 2010).

In Table VII, we examine the effect of education homophily on matching in the student teams. The effect of homophily in education is relatively weaker than gender and ethnicity. Attending the same undergraduate institution increases the probability of matching by 0.976%, while the homophily in gender and ethnicity are 1.33% and 1.38% respectively. In column 2 and 3, we observe the effect of homophily is much stronger among students from non-ivy league schools which typically have a lower representation among the overall student population. While attending the same college increases the matching probability by 1.88% among non-ivy school graduate, it only increases the matching rate by 0.219% among ivy-school graduates, despite the fact that there are far more Ivy graduates who attend HBS. It is important to note that the group size is much larger for

Ivy-league graduates. 24% of students are from the eight Ivy-league schools. The remaining 76% of students are from 85 non-Ivy league schools and each school represents less than 1% of the student population.

Table VIII explores the effect of school tie among male and female students. Brashears (2008) finds that homophily in education level is uniform among males and females using the data from 1985 general social survey. Our results point to a different story. The effect of a school tie is much stronger among males than it is among females. A school tie increases the matching rate by 1.71% among male students while it only increases matching rate by 0.096% among female students. Further, it is the strongest among male students from non-Ivy league schools. The difference with Brashears (2008) could be caused by difference in the setting that we examine, as Brashears (2008) examines education homophily in American's core discussion groups, people with whom Americans discuss "important matters". Another possible reason for the inconsistency is that Brashears (2008) uses educational level as the source for homophily, but we use educational institutions. It could be the case that while male and female both prefer to interact with people with the same educational level, male students cares more about the schools that others attended.

4.5 Past Industry Experience Homophily

Similarity in work experience can also be a source of homophily because it provides a common basis for socialization and friendships. On the other hand, teams may desire functional diversity as a way to improve performance, thus one might also expect that students who are seeking broad sets of skills may form teams with diverse work history (Ruef, Aldrich and Carter 2003). Table XV reports the results on industry matching. Our results show that at least in the context of the microbusiness formed as a part of Field 3, functional diversity was not an organizing principal. Industry homophily increases the probability of matching by 0.637%. Breaking down the homophily by industry sectors, we find homophily is strongest among people who worked in non-finance, consulting, or technology industry, and it increases the matching rate by 2.12%. The magnitude of the effects is similar among finance, technology, and consulting industries, which is around 0.35%.

Table X investigates the effect of industry homophily among male and female students. Male students are more likely to form groups with people who have the same industry experience. Industry tie increases the probability of matching by 0.887% among males. This is primarily driven by male students with experience in finance and technology. In contrast, industry tie only increases the matching rate by 0.292% among females. Interestingly, homophily is stronger among females with consulting background. It is important to note the stark contrast of homophily effect between males and females from finance industry. Male students with finance background are 0.8% (1% statistical significance) more likely to form teams with people from finance industry, but the industry homophily does not exist among females with finance backgrounds.

5. Homophily and Performance

The results in the previous section demonstrates that when students are allowed to choose their own teams to start a microbusiness, the propensity to pair up is increased by common personal characteristics including ethnicity, gender, education, and work experience. The other important aspect of our data is that the teams were assigned for the Class of 2013. We can therefore look at the causal relationship between performance and diversity. In this section, we examine the effect of homophily on team performance. While we look at the results for all classes, the results for the Classes of 2014-2016 need to be viewed with caution because of the endogeneity of group diversity.

In order to examine performance, our unit of observation is now team. There are 150-180 teams in each class year, and each team has 5-7 students. We measure team homophily across four different dimensions: Ethnicity, Gender, Education and Industry, and construct the homophily measure for each dimension as the following:

$$Homophily\ Score_{\ i} = \frac{\#\ of\ ties\ between\ team\ members\ with\ the\ same\ characterisitcs}{Total\ possible\ ties\ in\ the\ team}$$

To illustrate our homophily measure, consider a team with six people: Three of them are white, two of them are Asian Americans, and one is an international student from South America. "Ethnicity Score" in this team will be (3+1)/(5+4+3+2+1)=4/15, as there are three ties between three white team members[9], one tie between two Asian American students and fifteen possible ties between six team members. "Homophily score" is increasing in team homophily. It equals to zero if everyone in the team has different characteristics and equals to one if everyone is the same type.

Table XI provides summary statistics on team homophily by year. The average "Ethnicity Score" from 2014 to 2016 is 0.281, implying on average, there are 3 to 4 students with the same ethnic background in a team of 6 people. The standard deviation is also high (0.216), suggesting the existence of highly diverse teams and highly homogenous teams. The benchmark measure is the "ethnicity score" of the entire section. Compared to the benchmark, the "ethnicity score" is 19% higher than the ethnicity score of the entire section from 2014 to 2016, while it is roughly equals to the benchmark in 2013. Further, we observe the increasing incidents of teams with all White American students. The average number of teams with all White American students is 1 in 2013, and it increases to 4.7 after 2013 (Online Appendix Table 4).

The average "Gender Score" from 2014 to 2016 is 0.574, implying 4 to 5 people having the same gender in a team of 6. Comparing to the benchmark, "Gender Score" is 12.38% higher in

[9] When counting the tie between white people, we count Canadians (3.41%) and Australians (1.26%) also as white Americans. For other international students, we categorize their home countries by regions: Europe (8.8%), South Asia (6.9%), East Asia (4.6%), Latin America (4.6%), Middle East (3.8%), Africa (1.5%). A homophilous tie is recorded if two international students are from the same region (Online Appendix Table I).

2014-2016, and it is lower than the benchmark in 2013, reflecting the team assignment scheme utilized by the MBA administration. In addition, in 2013, there are no teams with all males or all female members. From 2014 to 2016, there are 20 teams with all male members and 8.3 teams with all female members per year[10].

The average "School Score" is 0.018. Approximately 1 out of 4 teams will have a pair of students from the same school. The "School Score" is 20% higher than the section benchmark from 2014 to 2016, while it is 5% lower than the benchmark in 2013. It is interesting that the benchmark of "School Score" is much higher in 2013. This may due to higher proportion of top college graduates (41.2%) in 2013 compared to 2014 to 2016 (37%)[11]. The average "Industry Score" is 0.21, implying around 3 people have the same industry background in a team of 6. The "Industry Score" is 8% higher than the benchmark in 2014-2016. Comparing class year of 2013 to 2014-106, homophily increases in all four dimensions in the 2014-2016 cohorts.

The HBS MBA Program office also provided the outcome of each team's microbusiness and we coded the outcome into four binary indicators: (1) "IPO Day": whether the team presents on the "IPO Day". Approximately 75% of the teams were deemed good enough to present on the "IPO Day"; (2) "Viable": whether the team that presented on the IPO day was deemed by judges to be viable. Roughly 50% of all projects are deemed "viable"; (3) "Section Top 3": whether the project was ranked in the top 3 of their section by the judges. Approximately 20% of the projects are "section top 3"; (4) "Class Top 3": whether the project is top 3 in the entire class year (2%).

We construct our performance measure based upon the median of the quantile of the team's project outcome. If a team does not present on the "IPO Day", the performance equals to 0.125, i.e., 25% of teams do not present, hence the median of this quantile is 0.125. Similarly, if a team

[10] See Online Appendix Table 4

[11] See Table I

presents on the "IPO Day" but the project is deemed not "viable", the performance equals to 0.375. The quantile in which this project performs falls between 25% and 50% of the class. Projects that are deemed viable but are not top 3 in the section have performance equal to 0.65, as their quantile falls between 50% and 80%. Projects that are top 3 in the section but not in the class year have performance equal to 0.9, i.e, falling between 80% and 98%. Finally, if the project is top 3 in the entire class year, the performance is 0.99. Our performance measure is increasing in project outcome. The distribution of "performance" does not vary significantly by year.

Panel C of the Table XI provides correlation table between variables. From 2014 to 2016, years in which matching is voluntary, we observe highly positive correlation between team "ethnicity score" and "school score", this is driven by White Americans and Asian Americans who attend top colleges. The correlation between "gender score" and "industry score" is also high, this may due to high percentage of male students with finance and technology industry experience. In 2013, in which the matching is forced by school, "school score" and "industry score" have slightly negative correlation with "ethnicity score" and "gender score". Interestingly, there is high correlation between "gender score" and "ethnicity score" in 2013. Further, "ethnicity score" is highly correlated with performance both in 2013 and in 2014-2016.

We split the sample into 2013 teams and 2014-2016 teams and run OLS regression on each sample. Because the team assignments in 2013 are assigned by the school, it provides a clean identification of the effect of homophily on performance. We estimate the following regression models:

$$Performance_i = b_{11} * \ Ethnicity \ Score_i + control + e_i$$
$$Performance_i = b_{21} * \ Gender \ Score_i + control + e_i$$
$$Performance_i = b_{31} * \ Education \ Score_i + control + e_i$$
$$Performance_i = b_{41} * \ Industry \ Score_i + control + e_i$$

Our performance measures are the median quantile of the team's project ranking, and our ethnicity (gender, education or industry) score is homophily measure which increases in team homophily. Control variables include team size, percentage of students who graduated from a top college and percentage of students who had start-up experience. Top college and start-up experience are proxies for students' ability and we expect these two variables to be positively correlated with performance.

Table XII reports the regression result of homophily on performance. Panel A column 1 shows that one unit increase in our "ethnicity score" increases team performance by 0.482 ($p<1\%$), or equivalently, one standard deviation increase in "ethnicity score" increases performance by 0.084[12]. Given the average performance of all teams is 0.5, this represents a 16.8% increase in performance. In addition to ethnicity, homophily in education is also positively correlated with performance. One unit increases in school score increases performance by 92.5%. In standard deviation term, one standard deviation increase in "school score" increases team performance by 0.027[13], 5.4% increases from the average performance. In column 6 of the panel A, the statistical significance on school score drops when we control for percentage of students who graduated from a top school and students who had start-up experience, indicating a positive correlation between school score and students' skill. Homophily in gender and past industry experience are positively correlated with performance, but the coefficients are not statistically significant. As a robustness test, we also use excess homophily score, defined as homophily score minus benchmark, as independent variables in Online Appendix Table 6. Our results are qualitatively identical with this adjustment. Panel B of Table XII reports the results of performance regression using 2014-2016 sample, where

12 We simulated the distribution of "ethnicity score" under the assumption of random matching. The SD of "ethnicity score" is 0.174 (Online Appendix Table 6, Panel C). 0.174* 0.482= 0.084

13 The SD of "school score" is 0.029 (Online Appendix Table 6, Panel C). 0.029*0.925= 0.027

the team formation is voluntary. The coefficient on ethnicity score is still positive and significant, but the magnitude is less than half of the 2013 result. Interestingly, the coefficients on "school score" and "industry score" reverses sign, implying homophily in education and industry is negatively correlated with team performance in endogenously formed teams, but the coefficients are not statistically significant.

In Table XIII, we investigate the cause of this performance change. We divide our performance measure into four dummy variables: ipo_day, viable, section top 3 and class year top 3. The results in panel A show that "ethnic score" increases the probability of presenting on IPO day and the project being judged viable, but it does not have as significant impact on the team's chance of being top three in their section or class year. Thus, greater homogeneity of ethnicity can help the worst performing teams to become average teams, but it has little effect on inducing superior performance. In other words, it avoids the worst outcomes. One possible explanation is that ethnically homogenous teams spend more time on projects as students receive positive utility from working with each other. Similarly, disagreements may be less likely in these groups as communication may be easier. As a result, they are less likely to be at the bottom of the performance spectrum, but they do not generate outcomes in the right hand tail of the distribution.

While we find greater ethnic homogeneity is positively correlated with performance, we do not conclude that diversity decreases performance or is undesirable for the following two reasons. First, our results on ethnicity lack generalizability beyond the range of diversity we see in our sample. Because the assignment of teams was done by the School with the intent of having relatively diverse teams, most teams in 2013 were relatively diverse in terms of gender and ethnic composition. In Figure 2, the graph plots team performance against "ethnicity score". The ethnicity score for most of teams falls below 50%, with mean equals to 23.7%. This implies that, on average, a team of 6 is comprised of students from 3-4 different ethnic groups[14]. Since there are too few extremely homogenous teams, we are not able to draw a conclusion on the effect of ethnic diversity across the entire spectrum of diversity. Similarly, Figure 3 shows that "gender score" for the Class of 2013 cohort concentrates between 40%-45% and has very little variation. Second, "assigned team diversity" does not guarantee harmony, as biases may still exist within the team. In fact, research has shown that mandatory diversity training actually reduces diversity in organizations (Dobbin and Kalev 2016; Dobbin, Kalev and Kelly 2007).

6. Conclusion and Discussion

In this paper, we examine the effect of homophily on entrepreneurial team formation using a unique dataset of MBA students. We also investigate the causal relationship between diversity and team performance. Our findings can be summarized as the followings.

The strength of demographic homophily (gender and ethnicity) is much stronger than homophily based upon acquired characteristics (education and industry). Specifically, homophily in gender and ethnicity increases the probability of matching by 25%. Homophily in education and past industry experience increases the probability of matching by 17% and 11% respectively.

Homophily is often stronger among smaller groups. Homophily is strongest among international students from the same region (Tale VII), students who attend non-Ivy league schools (Table XII) and students who worked in non-finance, consulting, or technology industries (small industries) (Table XV). One exception is Hispanic American students. Homophily among Hispanic American students is close to zero (Table VII).

[14] For a team of 6, if 3 people are white, 1 person is Asian American, 1 person is Hispanic American and another person is from Europe, this team will have ethnicity score = 3/15=20%

The effect of homophily in education and past industry experience is different for males and females. Males exhibit more propensity to match with people from the same school or who have worked in the same industry, particularly in technology and finance industry (Table XVI, XIII).

Homophily in ethnicity increases team performance by lifting teams in bottom quantiles to median performance quantiles, probably because it reduces conflicts and enhance communication efficiency within the group, but it does not induce superior performance in the right hand tail of the distribution.

Our results have important real-world implications given a significant portion of the MBA students will be working in the start-ups and venture capital industry. First, documenting the relative strength of the forces that cause people to associate sheds light on which factors are critical for limiting diversity in organizations like venture capital and entrepreneurship. To the extent that we observe the significant effect of various measures of homophily among MBA students, it is reasonable to infer that such homophily also exists in start-up team formation, venture capital investing, and hiring. If one goal of research is to identify the primary drivers that limit diversity, understanding the relative contribution of various factors is critical.

Second, in order to bring diversity into entrepreneurship, one needs to think carefully about how subtle treatment effects may dislodge the biases that occur based upon homophily in social networks. Results for the gender of venture capitalists' children (Gompers and Wang, 2017) show that when venture capitalists have more daughters, they are more likely to hire a female investor. Other subtle treatment effects may also debias individuals towards team homogeneity in terms of ethnicity as well. Our hope is the more research can explore the effectiveness of such subtle treatment effects for promoting greater organizational diversity.

We also find that ethnic homogeneity reduces the likelihood of bad outcomes, but does not increase the likelihood of extremely positive outcome. We caveat this result by noting that extreme homogeneity may actually reduce overall performance given that we do not observe extremely homogenous teams in the Class of 2013. We can, however, say that in industries like venture capital and startups, which make all of their returns on outcome in the right hand tail of the performance distribution, homogeneity does not appear to improve outcomes.

Finally, in this paper, while we document that demographic homophily is stronger than homophily in acquired characteristics, we do not attempt to trace the source of homophily. There are different views on why homophily exists in the economics literature. One view is that homophily is in an agent's preference function (Jackson, 2014). Another view is that homophily is the result of agents' strategic decisions to reduce uncertainty (Kets and Sandroni, 2016). Presumably, homophily that arises from these two different motivations may have different implications on the team formation process and performance. We do not, however, distinguish the motivation behind homophily. Additional research in this area is also warranted and important to answering these critical questions.

Reference

Alesina, A., & Ferrara, E. L. (2005). Ethnic diversity and economic performance. *Journal of economic literature*, 43(3), 762-800.

Block, P., & Grund, T. (2014). Multidimensional homophily in friendship networks. *Network Science*, *2*(02), 189-212.

Brashears, M. E. (2008). Gender and homophily: Differences in male and female association in Blau space. *Social Science Research*, *37*(2), 400-415.

Cohen, L., Frazzini, A., & Malloy, C. (2010). Sell‐side school ties. *The Journal of Finance*, 65(4), 1409-1437.

Currarini, S., Jackson, M. O., & Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4), 1003-1045.

Currarini, S., Jackson, M. O., & Pin, P. (2010). Identifying the roles of race-based choice and chance in high school friendship network formation. *Proceedings of the National Academy of Sciences*, *107*(11), 4857-4861.

Currarini, S., Matheson, J., & Vega-Redondo, F. (2016). A simple model of homophily in social networks. *European Economic Review*, 90, 18-39.

Dobbin, F., & Kalev, A. (2016). Why diversity programs fail. Harvard business review, 94(7), 14.

Dobbin, F., Kalev, A., & Kelly, E. (2007). Diversity management in corporate America. Contexts, 6(4), 21-27.

Fiore, A. T., & Donath, J. S. (2005, April). Homophily in online dating: when do you like someone like yourself?. *In CHI'05 Extended Abstracts on Human Factors in Computing Systems* (pp. 1371-1374). ACM.

Gompers, P. A., Mukharlyamov, V., & Xuan, Y. (2016). The cost of friendship. *Journal of Financial Economics*, 119(3), 626-644.

Gompers, P. A., and Wang, S.Q. (2017). Diversity in Innovation. Harvard Business School Working Paper, No. 17-067, January 2017.

Gompers, P. A., and Wang, S.Q. (2017). And the Children Shall Lead: The Effect of Children's Gender on Hiring and Performance in Venture Capital. Working Paper

Hampton, K. N., & Wellman, B. (2000). Examining community in the digital neighborhood: early results from Canada's wired suburb. In *Digital cities* (pp. 194-208). Springer Berlin Heidelberg.

Hoogendoorn, S., & Van Praag, M. (2012). Ethnic diversity and team performance: a field experiment. *Working Paper.*

Jackson, M. O. (2014). Networks in the understanding of economic behaviors. *Journal of Economic Perspectives* 28, 3–22.

Kalmijn M. 1998. Intermarriage and homogamy: causes, patterns and trends. *Annu. Rev. Sociol.* 24:395–421

Kets, W., & Sandroni, A. (2016). A belief-based theory of homophily. Working Paper.

Lazear, E. P. (1999). Globalisation and the market for team-mates. *The Economic Journal*, 109(454), 15-40.

Lazarsfeld, P. F., & Merton, R. K. (1954). Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1), 18-66.

Louch, H. (2000). Personal network integration: transitivity and homophily in strong-tie relations. *Social networks*, *22*(1), 45-64.

Marsden, P. V. (1987). Core discussion networks of Americans. *American sociological review*, 122-131.

Marsden, P. V. (1988). Homogeneity in confiding relations. *Social networks*, *10*(1), 57-76.

Marx, B., Pons, V., & Suri, T. (2015). Diversity and team performance in a kenyan organization. Working paper.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, *27*(1), 415-444.

Neckerman, H. J. (1996). The stability of social groups in childhood and adolescence: the role of the classroom social environment. *Social Development*, *5*(2), 131-145.

Ruef, M., Aldrich, H. E., & Carter, N. M. (2003). The structure of founding teams: Homophily, strong ties, and isolation among US entrepreneurs. *American sociological review*, 195-222.

Sorenson, O., & Stuart, T. E. (2001). Syndication networks and the spatial distribution of venture capital investments 1. *American journal of sociology*, 106(6), 1546-1588.

Van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annu. Rev. Psychol.*, 58, 515-541.

Verbrugge, L. M. (1977). The structure of adult friendship choices. *Social forces*, 576-597.

Wimmer, A., & Lewis, K. (2010). Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook 1. *American Journal of Sociology*, 116(2), 583-642.

Figure 1. Probability of Matching Conditional On Same Ethnicity, Gender, School, Industry

This figure plots the probability of a student being matched to another student with same ethnicity, gender, school or industry background. In 2013, the matching is randomized by the school. From 2014 to 2016, the matching process is initiated by students.

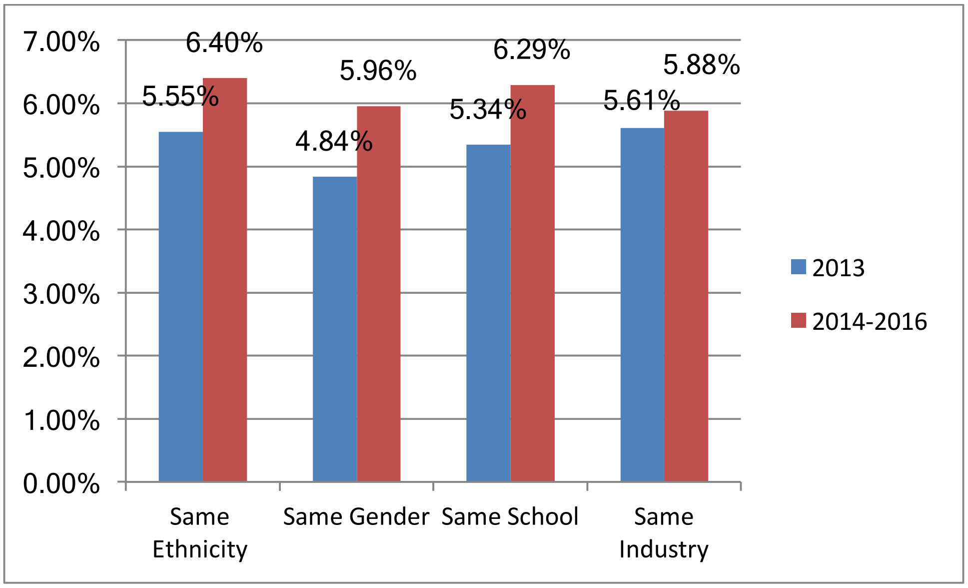


Table I. Summary Statistics of MBA Backgrounds

Table I presents the summary statistics of HBS MBA background from 2013 to 2016.

	2013	2014	2015	2016	Total
# of Students	907	915	931	931	3684
Team Size	6.06	6.13	6.25	5.2	5.91
Age	28.89	29.1	29.07	29.21	29.06
% of Female	39.25%	40.44%	41.14%	41.35%	40.55%
% of White American	37.16%	39.45%	37.70%	39.53%	38.46%
% of Asian American	14.33%	11.80%	11.92%	11.82%	12.46%
% of African American	4.52%	5.68%	5.59%	5.80%	5.40%
% of Hispanics American	3.75%	4.26%	4.83%	3.65%	4.13%
% International	34.07%	34.32%	34.59%	37.06%	35.02%
Employment Background					
% Finance Background	29.66%	29.29%	33.83%	36.84%	32.44%
% Consulting Background	21.94%	20.55%	20.62%	25.13%	22.07%
% Technology Background	9.04%	9.84%	10.85%	13.96%	10.94%
% Healthcare Background	8.16%	7.87%	6.34%	8.92%	7.82%
Education Background					
% Ivy League	26.90%	25.03%	23.63%	22.99%	24.62%
% Top School	41.23%	37.92%	38.35%	34.26%	37.92%

Table II. Past Employment and Education Background

This table summarizes the employment and education background of HBS MBAs.

Rank	Company	Obs	Percent
1	McKinsey & Company	308	8.40%
2	Bain & Company	184	5.02%
3	Boston Consulting Group	173	4.72%
4	Goldman Sachs	166	4.53%
5	Morgan Stanley	138	3.77%
6	Google	78	2.13%
7	Credit Suisse	54	1.47%
8	J.P. Morgan	47	1.28%
9	Deloitte Consulting	45	1.23%
10	Booz & Company	44	1.20%
11	UBS Investment Bank	42	1.15%
12	Bank of America Merrill Lynch	38	1.04%
13	Bain Capital	32	0.87%
14	United States Marine Corps	29	0.79%
15	Accenture	26	0.71%
15	Citigroup	26	0.71%
15	Barclays Capital	25	0.68%
15	Oliver Wyman	25	0.68%
15	The Blackstone Group	25	0.68%
20	Deutsche Bank	24	0.65%
20	The Carlyle Group	24	0.65%
	Top 20 Total	1553	42.37%
	Sample Total	3,665	

Rank	School	Obs	Percent
1	Harvard University	286	8.17%
2	Stanford University	157	4.49%
3	University of Pennsylvania	151	4.31%
4	Yale University	124	3.54%
5	Princeton University	102	2.91%
6	Duke University	81	2.31%
7	MIT	72	2.06%
8	United States Military Academy	70	2.00%
9	Dartmouth College	67	1.91%
10	University of California	64	1.83%
11	Cornell University	63	1.80%
12	Georgetown University	60	1.71%
13	Brown University	57	1.63%
13	Columbia University	57	1.63%
15	Northwestern University	56	1.60%
16	University of Virginia	52	1.49%
17	Indian Institute of Technology	50	1.43%
18	University of Texas	45	1.29%
19	University of Michigan	38	1.09%
20	Brigham Young University	37	1.06%
	Top 20 Total	1689	48.26%
	Sample Total	3,500	

Table III. Matching Regression

This table reports the regression results of matching on ethnicity (gender, education, industry) ties. Each observation is a student-student pair. The dependent variable "real_match" equals to 1 if the pair is in the same team. The independent variables "ethnicity (gender, education, industry) match" equals to 1 if the pair has the same ethnicity (gender, education, industry).

Panel A. 2014-2016	(1)	(2)	(3)	(4)	(5)
VARIABLES	Real Match	Real Match	Real Match	Real Match	Real Match
Ethnicity Tie	0.0138***				0.0136***
	(0.00116)				(0.00116)
Gender Tie		0.0133***			0.0131***
		(0.00107)			(0.00106)
School Tie			0.00976**		0.00855**
			(0.00384)		(0.00383)
Industry Tie				0.00637***	0.00625***
				(0.00120)	(0.00120)
Team Mem Count	0.0106***	0.0109***	0.0108***	0.0107***	0.0105***
	(0.000114)	(5.65e-05)	(2.42e-05)	(5.04e-05)	(0.000132)
2015.ClassYear	-0.000746***	-0.000957***	-0.000983***	-0.00116***	-0.000895***
	(0.000123)	(6.21e-05)	(2.52e-05)	(5.48e-05)	(0.000145)
2016.ClassYear	-0.00121***	-0.000945***	-0.00105***	-0.00167***	-0.00174***
	(0.000162)	(7.82e-05)	(3.07e-05)	(0.000134)	(0.000223)
Constant	-0.0112***	-0.0167***	-0.00958***	-0.00981***	-0.0186***
	(0.000692)	(0.000680)	(0.000148)	(0.000273)	(0.000974)
Observations	254,318	254,318	254,318	254,318	254,318
R-squared	0.002	0.002	0.001	0.001	0.003

Panel B.2013	(6)	(7)	(8)	(9)	(10)
VARIABLES	Real Match	Real Match	Real Match	Real Match	Real Match
Ethnicity Tie	-0.00116				-0.000837
	(0.00170)				(0.00170)
Gender Tie		-0.0166***			-0.0166***
		(0.000716)			(0.000716)
School Tie			-0.00303		-0.00284
			(0.00604)		(0.00605)
Industry Tie				-0.000367	-0.000271
				(0.00215)	(0.00215)
Team Mem Count	0.0106***	0.0106***	0.0106***	0.0106***	0.0106***
	(5.93e-05)	(0.000220)	(5.49e-05)	(5.56e-05)	(0.000218)
Constant	-0.00772***	0.000630	-0.00788***	-0.00789***	0.000883
	(0.000490)	(0.00138)	(0.000359)	(0.000454)	(0.00143)
Observations	81,368	81,368	81,368	81,368	81,368
R-squared	0.000	0.001	0.000	0.000	0.001

Table IV. Ethnicity Match Regression

This table reports the regression results of the probability of match on ethnicity ties. Each observation is a student-student pair. The dependent variable real match equals to 1 if the students are teammates. The independent variables are ethnicity characteristics equals to 1 if both students share the same ethnicity.

2014-2016	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Real Match	Real Match	Real Match	Real Match	Real Match	Real Match
Both White	0.0108***					0.0118***
	(0.00114)					(0.00116)
Both Asian American		0.0116***				0.0145***
		(0.00421)				(0.00423)
Both African American			0.00956			0.0126
			(0.00869)			(0.00870)
Both Hispanic American				5.47e-05		0.00306
				(0.0125)		(0.0125)
Both International					0.0377***	0.0401***
					(0.00506)	(0.00508)
Team Mem Count	0.0106***	0.0108***	0.0108***	0.0108***	0.0109***	0.0107***
	(0.000104)	(2.93e-05)	(2.32e-05)	(2.37e-05)	(4.89e-05)	(9.12e-05)
2015.ClassYear	-0.000773***	-0.000982***	-0.000983***	-0.000983***	-0.00105***	-0.000827***
	(0.000111)	(3.18e-05)	(2.38e-05)	(2.42e-05)	(5.33e-05)	(9.79e-05)
2016.ClassYear	-0.00121***	-0.00105***	-0.00104***	-0.00104***	-0.000966***	-0.00114***
	(0.000146)	(3.75e-05)	(2.85e-05)	(3.03e-05)	(6.73e-05)	(0.000126)
Constant	-0.0103***	-0.00960***	-0.00957***	-0.00951***	-0.0106***	-0.0118***
	(0.000613)	(0.000179)	(0.000151)	(0.000153)	(0.000325)	(0.000564)
Observations	254,318	254,318	254,318	254,318	254,318	254,318
R-squared	0.001	0.001	0.001	0.001	0.001	0.002

2013	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	Real Match	Real Match	Real Match	Real Match	Real Match	Real Match
Both White	1.17e-05					-0.000239
	(0.00174)					(0.00177)
Both Asian American		0.00222				0.00193
		(0.00508)				(0.00512)
Both African American			0.000214			-3.45e-05
			(0.0182)			(0.0182)
Both Hispanic American				0.00439		0.00414
				(0.0219)		(0.0219)
Both International					-0.0158***	-0.0158***
					(0.00530)	(0.00534)
Team Mem Count	0.0106***	0.0106***	0.0106***	0.0106***	0.0106***	0.0106***
	(6.02e-05)	(8.38e-05)	(7.71e-05)	(5.69e-05)	(7.45e-05)	(0.000117)
Constant	-0.00794***	-0.00816***	-0.00794***	-0.00793***	-0.00750***	-0.00766***
	(0.000372)	(0.000591)	(0.000451)	(0.000345)	(0.000478)	(0.000795)
Observations	81,368	81,368	81,368	81,368	81,368	81,368
R-squared	0.000	0.000	0.000	0.000	0.000	0.000

Table V. Gender Match Regression

This table reports the regression results of the probability of match on Gender ties. Each observation is a student-student pair. The dependent variable real match equals to 1 if the students are teammates. The independent variables are *Both Male (Female)* equals to 1 if both students are male (female).

	2014-2016	2014-2016	2013	2013
	(1)	(2)	(3)	(4)
VARIABLES	Real Match	Real Match	Real Match	Real Match
Both Male	0.00723***		-0.00897***	
	(0.000868)		(0.000596)	
Both Female		0.0122***		-0.0161***
		(0.00130)		(0.000897)
Team Mem Count	0.0109***	0.0107***	0.0106***	0.0106***
	(9.28e-05)	(0.000106)	(0.000320)	(0.000370)
2015.ClassYear	-0.000934***	-0.00104***		
	(0.000101)	(0.000115)		
2016.ClassYear	-0.000880***	-0.00123***		
	(0.000131)	(0.000151)		
Constant	-0.0126***	-0.0109***	-0.00468**	-0.00550**
	(0.000680)	(0.000654)	(0.00195)	(0.00225)
Observations	254,318	254,318	81,368	81,368
R-squared	0.001	0.001	0.001	0.001

Table VI. Gender Match Breakdown by Ethnicity

This table reports the regression results of the probability of match on Gender and ethnicity ties. Each observation is a student-student pair. The dependent variable real match equals to 1 if the students are teammates. The independent variable is ethnicity tie. First two columns look at the matching results of male subsample, last two columns look at the female subsample.

	2014-2016			
	Male	Male	Female	Female
	(1)	(2)	(3)	(4)
VARIABLES	Real Match	Real Match	Real Match	Real Match
Ethnicity Tie	0.0154***		0.0114***	
	(0.00151)		(0.00180)	
Both White		0.0135***		0.00914***
		(0.00150)		(0.00180)
Both Asian American		0.0147**		0.0141***
		(0.00672)		(0.00543)
Both African American		0.00326		0.0245*
		(0.0107)		(0.0142)
Both Hispanic American		0.0159		-0.0249
		(0.0158)		(0.0186)
Both International		0.0427***		0.0363***
		(0.00650)		(0.00815)
Team Mem Count	0.0105***	0.0106***	0.0107***	0.0107***
	(0.000170)	(0.000142)	(0.000143)	(0.000106)
2015.ClassYear	-0.000759***	-0.000802***	-0.000742***	-0.000878***
	(0.000181)	(0.000146)	(0.000159)	(0.000126)
2016.ClassYear	-0.00112***	-0.00103***	-0.00132***	-0.00122***
	(0.000235)	(0.000192)	(0.000214)	(0.000160)
Constant	-0.0111***	-0.0118***	-0.0112***	-0.0116***
	(0.00101)	(0.000856)	(0.000896)	(0.000698)
Observations	150,093	150,093	104,225	104,225
R-squared	0.002	0.002	0.001	0.002

Table VII. Education Match Regression

This table reports the regression results of the probability of match on education ties. Each observation is a student-student pair. The dependent variable real match equals to 1 if the students are teammates. The independent variables *Both Same (Non) Ivy School* equals to 1 if both students are graduated from the same (Non) Ivy schools.

	2014-2016	2014-2016	2014-2016	2014-2016	2013	2013	2013	2013
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Real Match	Real Match	Real Match	Real Match	Real Match	Real Match	Real Match	Real Match
School Tie	0.00976**				-0.00303			
	(0.00384)				(0.00604)			
Both Ivy School		0.00219		0.00232		0.00623		0.00611
		(0.00490)		(0.00490)		(0.00852)		(0.00852)
Both Non Ivy			0.0188***	0.0189***			-0.0145*	-0.0144*
			(0.00600)	(0.00600)			(0.00821)	(0.00821)
Team Mem Count	0.0108***	0.0108***	0.0108***	0.0108***	0.0106***	0.0106***	0.0106***	0.0106***
	(2.42e-05)	(2.30e-05)	(2.38e-05)	(2.45e-05)	(5.49e-05)	(5.64e-05)	(6.04e-05)	(6.35e-05)
2015.ClassYear	-0.000983***	-0.000981***	-0.000996***	-0.000994***				
	(2.52e-05)	(2.37e-05)	(2.64e-05)	(2.65e-05)				
2016.ClassYear	-0.00105***	-0.00104***	-0.00104***	-0.00105***				
	(3.07e-05)	(2.83e-05)	(3.04e-05)	(3.08e-05)				
Constant	-0.00958***	-0.00952***	-0.00960***	-0.00961***	-0.00788***	-0.00803***	-0.00786***	-0.00795***
	(0.000148)	(0.000137)	(0.000148)	(0.000148)	(0.000359)	(0.000366)	(0.000372)	(0.000408)
Observations	254,318	254,318	254,318	254,318	81,368	81,368	81,368	81,368
R-squared	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000

Table VIII. Education Match Regression by Gender

This table reports the regression results of the probability of match on education ties by gender. Each observation is a student-student pair. The dependent variable real match equals to 1 if the students are teammates. The independent variables *Both Same (Non) Ivy School* equals to 1 if both students are graduated from the same (Non) Ivy schools.

	2014-2016			
	Male	Male	Female	Female
	(1)	(2)	(3)	(4)
VARIABLES	Real Match	Real Match	Real Match	Real Match
School Tie	0.0171***		0.000960	
	(0.00540)		(0.00540)	
Both Ivy School		0.00893		-0.00391
		(0.00714)		(0.00671)
Both Non Ivy		0.0250***		0.00898
		(0.00804)		(0.00889)
Team Mem Count	0.0108***	0.0108***	0.0108***	0.0108***
	(3.59e-05)	(3.55e-05)	(3.42e-05)	(3.65e-05)
2015.ClassYear	-0.000971***	-0.000969***	-0.000993***	-0.00102***
	(3.66e-05)	(3.65e-05)	(3.67e-05)	(4.40e-05)
2016.ClassYear	-0.00106***	-0.00106***	-0.00106***	-0.00105***
	(4.53e-05)	(4.42e-05)	(4.35e-05)	(4.68e-05)
Constant	-0.00961***	-0.00961***	-0.00948***	-0.00953***
	(0.000212)	(0.000209)	(0.000220)	(0.000231)
Observations	150,093	150,093	104,225	104,225
R-squared	0.001	0.001	0.001	0.001

Table IX. Past Employment Regression

This table reports the regression results of the probability of match on education ties. Each observation is a student-student pair. The dependent variable real match equals to 1 if the students are teammates. The independent variables are industry backgrounds equals to 1 if both students worked in the same industry prior to MBA.

2014-2016	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Real Match	Real Match	Real Match	Real Match	Real Match	Real Match
Industry Tie	0.00637***					
	(0.00120)					
Both Finance Industry		0.00346**				0.00418***
		(0.00142)				(0.00144)
Both Tech Industry			0.00362			0.00455
			(0.00378)			(0.00378)
Both Consulting Industry				0.00354*		0.00432**
				(0.00190)		(0.00191)
Both Small Industry					0.0212***	0.0218***
					(0.00391)	(0.00392)
Team Mem Count	0.0107***	0.0107***	0.0108***	0.0108***	0.0109***	0.0108***
	(5.04e-05)	(5.04e-05)	(2.36e-05)	(2.81e-05)	(3.59e-05)	(5.25e-05)
2015.ClassYear	-0.00116***	-0.00107***	-0.000991***	-0.000985***	-0.000964***	-0.00108***
	(5.48e-05)	(4.73e-05)	(2.57e-05)	(2.66e-05)	(3.55e-05)	(5.02e-05)
2016.ClassYear	-0.00167***	-0.00129***	-0.00108***	-0.00110***	-0.00112***	-0.00154***
	(0.000134)	(0.000113)	(5.04e-05)	(4.57e-05)	(5.00e-05)	(0.000132)
Constant	-0.00981***	-0.00929***	-0.00952***	-0.00978***	-0.0102***	-0.0103***
	(0.000273)	(0.000229)	(0.000141)	(0.000216)	(0.000255)	(0.000309)
Observations	254,318	254,318	254,318	254,318	254,318	254,318
R-squared	0.001	0.001	0.001	0.001	0.001	0.001
2013	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	Real Match	Real Match	Real Match	Real Match	Real Match	Real Match
Industry Tie	-0.000367					
	(0.00215)					
Both Finance Industry		-0.000909				-0.000839
		(0.00254)				(0.00258)
Both Tech Industry			0.0215**			0.0213**
			(0.0102)			(0.0103)
Both Consulting Industry				-0.00538		-0.00517
				(0.00380)		(0.00384)
Both Small Industry					0.00692	0.00681
					(0.00542)	(0.00544)
Team Mem Count	0.0106***	0.0106***	0.0106***	0.0106***	0.0106***	0.0107***
	(5.56e-05)	(5.97e-05)	(9.53e-05)	(7.12e-05)	(6.61e-05)	(0.000111)
Constant	-0.00789***	-0.00788***	-0.00818***	-0.00776***	-0.00817***	-0.00817***
	(0.000454)	(0.000402)	(0.000593)	(0.000451)	(0.000443)	(0.000759)
Observations	81,368	81,368	81,368	81,368	81,368	81,368
R-squared	0.000	0.000	0.000	0.000	0.000	0.000

Table X. Past Employment Regression by Gender

This table reports the regression results of the probability of match on education ties by gender. Each observation is a student-student pair. The dependent variable real match equals to 1 if the students are teammates. The independent variables are industry backgrounds equals to 1 if both students worked in the same industry prior to MBA.

	2014-2016			
	Male	Male	Female	Female
	(1)	(2)	(3)	(4)
VARIABLES	Real Match	Real Match	Real Match	Real Match
Industry Tie	0.00887***		0.00292	
	(0.00159)		(0.00181)	
Both Finance Industry		0.00801***		-0.00172
		(0.00188)		(0.00219)
Both Tech Industry		0.00869*		-0.00220
		(0.00484)		(0.00603)
Both Consulting Industry		0.00259		0.00564**
		(0.00278)		(0.00265)
Both Small Industry		0.0214***		0.0226***
		(0.00487)		(0.00659)
Team Mem Count	0.0106***	0.0107***	0.0108***	0.0109***
	(8.76e-05)	(9.53e-05)	(5.00e-05)	(7.81e-05)
2015.ClassYear	-0.00107***	-0.00108***	-0.00114***	-0.000823***
	(7.66e-05)	(6.68e-05)	(0.000104)	(0.000132)
2016.ClassYear	-0.00182***	-0.00178***	-0.00139***	-0.00113***
	(0.000179)	(0.000181)	(0.000212)	(0.000223)
Constant	-0.00976***	-0.00998***	-0.00966***	-0.0106***
	(0.000468)	(0.000533)	(0.000295)	(0.000523)
Observations	150,093	150,093	104,225	104,225
R-squared	0.001	0.001	0.001	0.001

Table XI. Summary Statistics on Team Homophily and Performance

This table reports the summary statistics on the team homophily scores and performance.

Panel A. Homophily Measures								
2013								
Variable	Obs	Mean	Benchmark	Mean/Benchmark	Std. Dev.	SE	Min	Max
Team Member Count	150	6.047			0.268	0.022	5.0	7.0
Ethnicity Score	150	0.237	0.242	98.01%	0.166	0.014	0.0	1.0
Gender Score	150	0.444	0.518	85.71%	0.038	0.003	0.4	0.7
School Score	150	0.017	0.018	95.75%	0.039	0.003	0.0	0.2
Industry Score	150	0.163	0.164	99.65%	0.136	0.011	0.0	0.9
2014								
Variable	Obs	Mean	Benchmark	Mean/Benchmark	Std. Dev.	SE	Min	Max
Team Member Count	150	6.100			0.414	0.034	5.0	7.0
Ethnicity Score	150	0.290	0.247	117.70%	0.214	0.017	0.0	1.0
Gender Score	150	0.582	0.513	113.46%	0.216	0.018	0.4	1.0
School Score	150	0.017	0.015	110.82%	0.035	0.003	0.0	0.2
Industry Score	150	0.181	0.153	118.26%	0.164	0.013	0.0	1.0
2015								
Variable	Obs	Mean	Benchmark	Mean/Benchmark	Std. Dev.	SE	Min	Max
Team Member Count	150	6.207			0.496	0.040	5.0	7.0
Ethnicity Score	150	0.271	0.232	116.89%	0.228	0.019	0.0	1.0
Gender Score	150	0.558	0.511	109.21%	0.202	0.017	0.4	1.0
School Score	150	0.019	0.016	115.87%	0.039	0.003	0.0	0.2
Industry Score	150	0.183	0.183	99.99%	0.145	0.012	0.0	0.7
2016								
Variable	Obs	Mean	Benchmark	Mean/Benchmark	Std. Dev.	SE	Min	Max
Team Member Count	180	5.172			0.393	0.029	4.0	6.0
Ethnicity Score	180	0.280	0.230	122.06%	0.241	0.018	0.0	1.0
Gender Score	180	0.582	0.510	114.11%	0.227	0.017	0.4	1.0
School Score	180	0.019	0.015	133.33%	0.047	0.004	0.0	0.3
Industry Score	180	0.255	0.235	108.59%	0.177	0.013	0.0	1.0
2014-2016 Average								
Variable	Obs	Mean	Benchmark	Mean/Benchmark	Std. Dev.	SE	Min	Max
Team Member Count	480	5.785			0.644	0.029	4.0	7.0
Ethnicity Score	480	0.281	0.236	119.04%	0.228	0.010	0.0	1.0
Gender Score	480	0.574	0.511	112.38%	0.216	0.010	0.4	1.0
School Score	480	0.018	0.015	120.60%	0.041	0.002	0.0	0.3
Industry Score	480	0.210	0.193	108.45%	0.167	0.008	0.0	1.0

Panel B. Performance Measures							
Class Year	Freq.	ipo year	viable	section top 3	classytop3	Performance	SD
2013	150	78.67%	46.67%	20.00%	2.67%	0.502	0.275
2014	150	70.00%	39.33%	20.00%	2.00%	0.460	0.290
2015	150	73.33%	55.33%	20.00%	2.00%	0.512	0.287
2016	180	76.11%	52.78%	16.67%	2.22%	0.504	0.272
Total	630	74.60%	48.73%	19.05%	2.22%	0.495	0.281

Panel C. Correlation Between Variables					
2014-2016	Ethnicity Score	Gender Score	School Score	Industry Score	Performance
Ethnicity Score	1				
Gender Score	-0.0262	1			
School Score	**0.1415**	-0.016	1		
Industry Score	0.0403	**0.1253**	0.0791	1	
Performance	**0.1556**	0.0203	-0.0042	-0.0355	1
2013	Ethnicity Score	Gender Score	School Score	Industry Score	Performance
Ethnicity Score	1				
Gender Score	**0.1013**	1			
School Score	-0.0166	-0.017	1		
Industry Score	-0.0371	-0.0819	0.0084	1	
Performance	**0.2907**	0.0324	**0.1303**	0.0309	1

Table XII. Homophily and Performance Regression

The dependent variable Performance=0.125 if the team does not present on IPO day (0-25%), =0.375 if present but not viable (25-50%), =0.65 if viable but not top 3 (50-80%), =0.9 if top 3 in section (80-98%), =0.99 if top 3 in class year (98-100%).

Panel A. 2013	[1]	[2]	[3]	[4]	[5]	[6]
VARIABLES				Performance		
Ethnicity Score	0.482***				0.488***	0.450***
	(0.139)				(0.148)	(0.148)
Gender Score		0.211			0.0249	0.0802
		(0.557)			(0.507)	(0.508)
School Score			0.925*		0.958*	0.809
			(0.556)		(0.535)	(0.593)
Industry Score				0.0635	0.0842	0.0719
				(0.148)	(0.152)	(0.159)
Top School Ratio						0.0845
						(0.111)
Start-up Ratio						0.529
						(0.386)
Team Mem Count	0.0471	0.0402	0.0491	0.0433	0.0538	0.0515
	(0.0634)	(0.0711)	(0.0687)	(0.0701)	(0.0609)	(0.0649)
Constant	0.103	0.166	0.190	0.230	0.0204	-0.0279
	(0.386)	(0.465)	(0.418)	(0.428)	(0.412)	(0.429)
Observations	150	150	150	150	150	150
R-squared	0.087	0.003	0.019	0.003	0.107	0.122

Panel B. 2014-2016	[1]	[2]	[3]	[4]	[5]	[6]
VARIABLES				Performance		
Ethnicity Score	0.176***				0.185***	0.168***
	(0.0536)				(0.0538)	(0.0540)
Gender Score		0.0273			0.0434	0.0641
		(0.0585)			(0.0582)	(0.0587)
School Score			-0.0915		-0.195	-0.298
			(0.297)		(0.292)	(0.298)
Industry Score				-0.110	-0.121	-0.113
				(0.0798)	(0.0812)	(0.0812)
Top School Ratio						0.0907
						(0.0558)
Start-up Ratio						0.341**
						(0.134)
Team Mem Count	0.0885***	0.0991***	0.0997***	0.105***	0.0942***	0.0847***
	(0.0277)	(0.0274)	(0.0275)	(0.0275)	(0.0277)	(0.0274)
2015.ClassYear	0.0462	0.0424	0.0419	0.0413	0.0473	0.0422
	(0.0329)	(0.0332)	(0.0333)	(0.0332)	(0.0328)	(0.0329)
2016.ClassYear	0.128***	0.136***	0.137***	0.149***	0.143***	0.127***
	(0.0390)	(0.0390)	(0.0391)	(0.0404)	(0.0405)	(0.0407)
Constant	-0.131	-0.161	-0.147	-0.158	-0.168	-0.163
	(0.168)	(0.171)	(0.168)	(0.168)	(0.170)	(0.167)
Observations	480	480	480	480	480	480
R-squared	0.049	0.030	0.030	0.033	0.056	0.074

Table XIII. Performance Breakdown

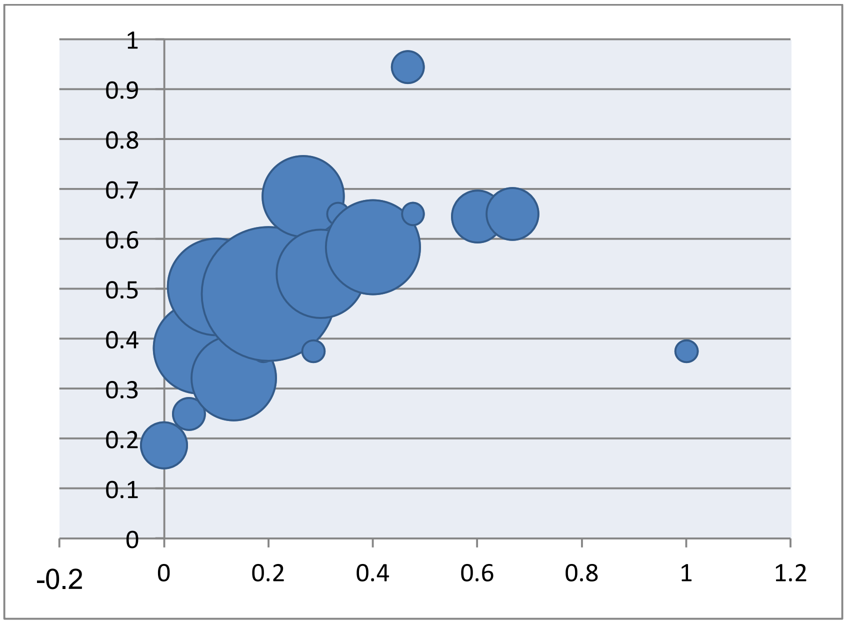
This table reports regression results using different performance measures. “ipo_day” equals to 1 if the team present on the ipo day. “viable” equals to 1 if the team presents on the ipo day and is deemed “viable” by the judges. “sect_top3” and “classytop3” are the top 3 in the section and class year.

Panel A. 2013	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
VARIABLES	ipo_day	ipo_day	viable	viable	sect_top3	sect_top3	classyrtop3	classyrtop3
Ethnicity Score	0.587***	0.541***	0.981***	0.923***	0.264	0.223	0.0610	0.0606
	(0.184)	(0.178)	(0.299)	(0.304)	(0.190)	(0.199)	(0.0747)	(0.0692)
Gender Score	0.570	0.664	-0.111	-0.0287	-0.500	-0.457	0.422*	0.403*
	(0.792)	(0.795)	(0.996)	(0.993)	(0.764)	(0.778)	(0.244)	(0.235)
School Score	0.997	0.658	1.066	0.849	1.405	1.343	0.722	0.834
	(0.680)	(0.785)	(1.024)	(1.114)	(0.964)	(1.019)	(0.514)	(0.550)
Industry Score	0.261	0.212	0.0853	0.0694	-0.00668	0.00261	-0.0333	-0.00802
	(0.231)	(0.244)	(0.309)	(0.323)	(0.209)	(0.220)	(0.0811)	(0.0897)
Top School Ratio		0.178		0.123		0.0437		-0.0536
		(0.189)		(0.211)		(0.151)		(0.0392)
Start-up Ratio		0.544		0.819		0.643		0.0787
		(0.483)		(0.711)		(0.651)		(0.262)
Team Mem Count	0.146	0.147	0.0864	0.0826	-0.0192	-0.0241	-0.0179	-0.0206
	(0.107)	(0.107)	(0.128)	(0.133)	(0.0954)	(0.0996)	(0.0110)	(0.0128)
Constant	-0.547	-0.662	-0.271	-0.340	0.453	0.436	-0.0738	-0.0337
	(0.713)	(0.711)	(0.861)	(0.886)	(0.616)	(0.640)	(0.0763)	(0.0720)
Observations	150	150	150	150	150	150	150	150
R-squared	0.083	0.095	0.113	0.124	0.032	0.041	0.047	0.052
Panel B. 2014-2016	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
VARIABLES	ipo_day	ipo_day	viable	viable	sect_top3	sect_top3	classyrtop3	classyrtop3
Ethnicity Score	0.266***	0.242***	0.265***	0.236**	0.180**	0.170**	0.000779	0.00124
	(0.0804)	(0.0805)	(0.0999)	(0.100)	(0.0796)	(0.0806)	(0.0295)	(0.0290)
Gender Score	0.0630	0.0932	0.0258	0.0615	0.0676	0.0811	0.0406	0.0400
	(0.0909)	(0.0928)	(0.103)	(0.103)	(0.0844)	(0.0849)	(0.0354)	(0.0373)
School Score	-0.0414	-0.238	-0.147	-0.328	-0.521	-0.526	-0.160	-0.187*
	(0.499)	(0.506)	(0.542)	(0.557)	(0.388)	(0.406)	(0.104)	(0.111)
Industry Score	-0.181	-0.174	-0.214	-0.201	-0.0778	-0.0672	0.0318	0.0290
	(0.136)	(0.135)	(0.141)	(0.140)	(0.100)	(0.101)	(0.0466)	(0.0464)
Top School Ratio		0.151*		0.158		0.0342		0.00966
		(0.0913)		(0.0989)		(0.0776)		(0.0215)
Start-up Ratio		0.314*		0.574**		0.466*		-0.126***
		(0.189)		(0.230)		(0.237)		(0.0419)
2015.ClassYear	0.0289	0.0249	0.146***	0.137**	-3.28e-06	-0.00771	-0.000267	0.00198
	(0.0519)	(0.0521)	(0.0564)	(0.0564)	(0.0464)	(0.0465)	(0.0166)	(0.0165)
2016.ClassYear	0.175***	0.160**	0.330***	0.304***	0.0287	0.00811	0.0137	0.0191
	(0.0649)	(0.0652)	(0.0702)	(0.0709)	(0.0591)	(0.0594)	(0.0242)	(0.0249)
Team Mem Count	0.105**	0.0935**	0.190***	0.174***	0.0572	0.0481	0.0144	0.0160
	(0.0454)	(0.0452)	(0.0488)	(0.0484)	(0.0399)	(0.0397)	(0.0172)	(0.0172)
Constant	-0.0225	-0.0285	-0.818***	-0.809***	-0.218	-0.195	-0.0946	-0.104
	(0.279)	(0.276)	(0.297)	(0.292)	(0.250)	(0.249)	(0.109)	(0.110)
Observations	480	480	480	480	480	480	480	480
R-squared	0.038	0.049	0.066	0.082	0.021	0.033	0.010	0.016

Figure 2. Performance and Ethnicity Score (2013, 2014-2016)

The Y axis is the performance of the team, X axis is the ethnicity score, ranges from 0 (most diverse) to 1 (homogenous). The size of the bubble is proportion to observation number.

2013 (Average Ethnicity Score=23.7%)



2014-2016 (Average Ethnicity Score=28.1%)

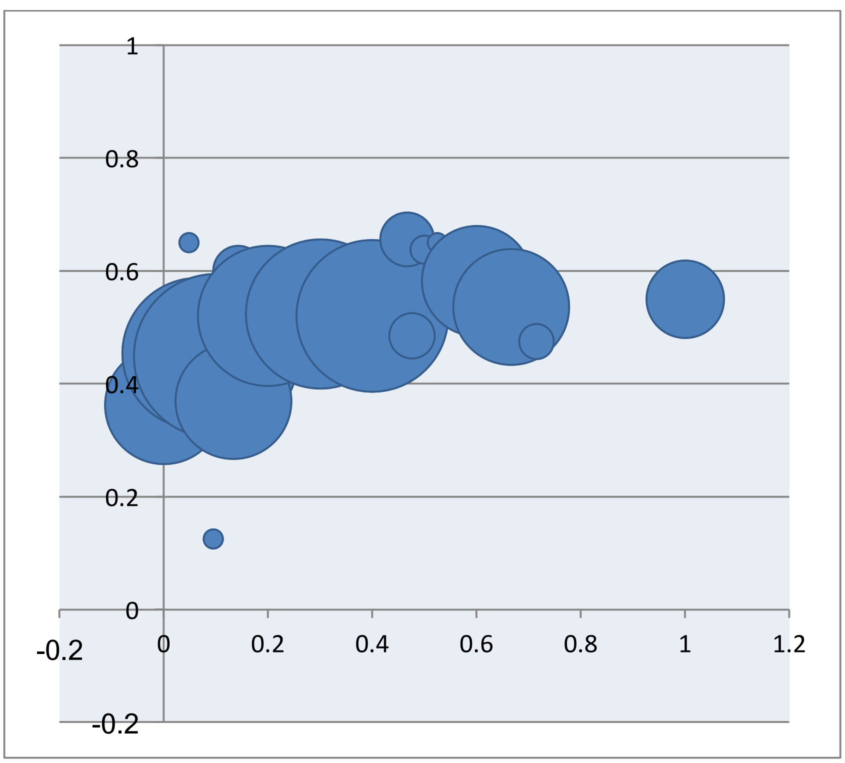
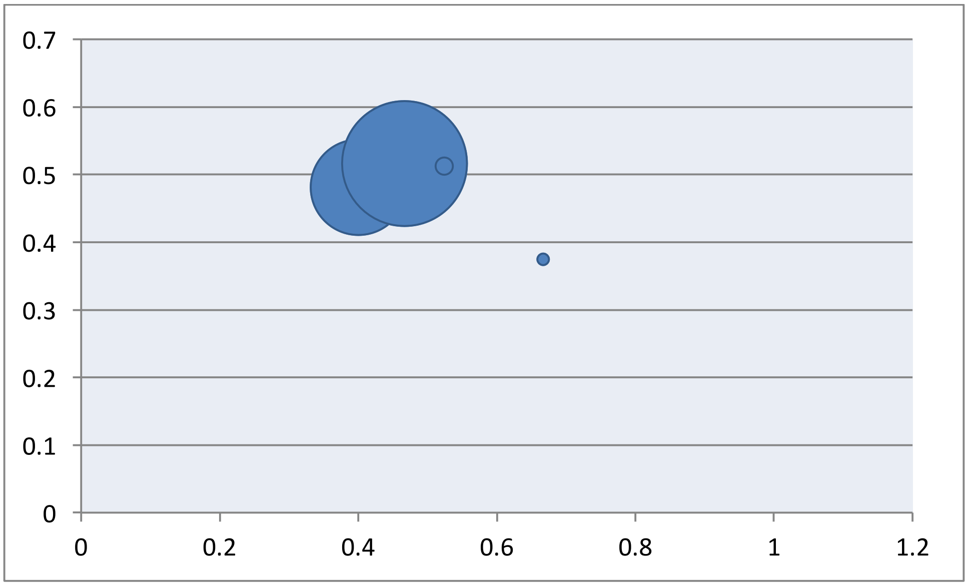


Figure 3. Performance and Gender Score (2013, 2014-2016)

The Y axis is the performance of the team, X axis is the gender score, ranges from 0 (most diverse) to 1 (homogenous). The size of the bubble is proportion to observation number.

2013 (Average Gender Score=44.4%)



2013 (Average Gender Score=57.4%)

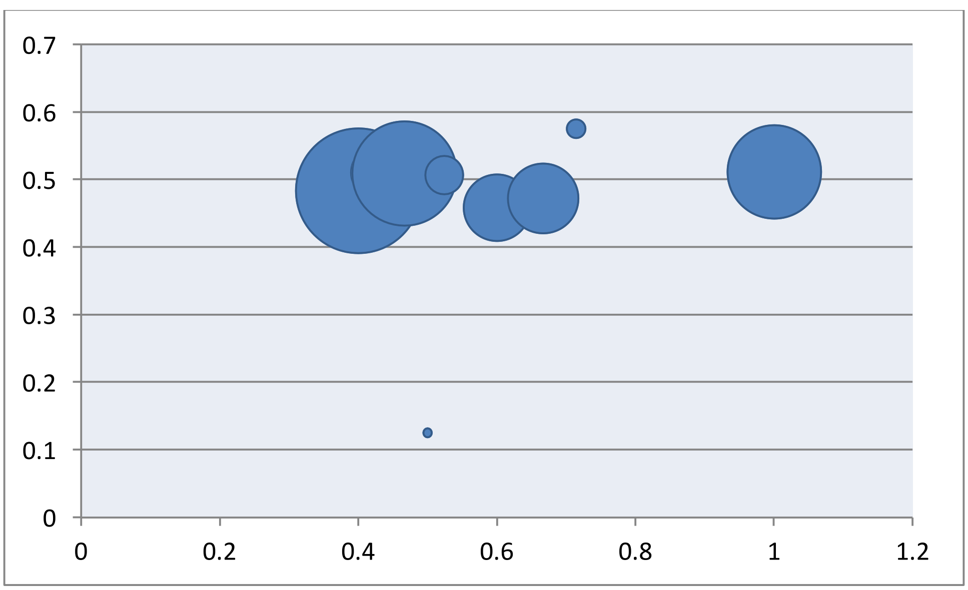
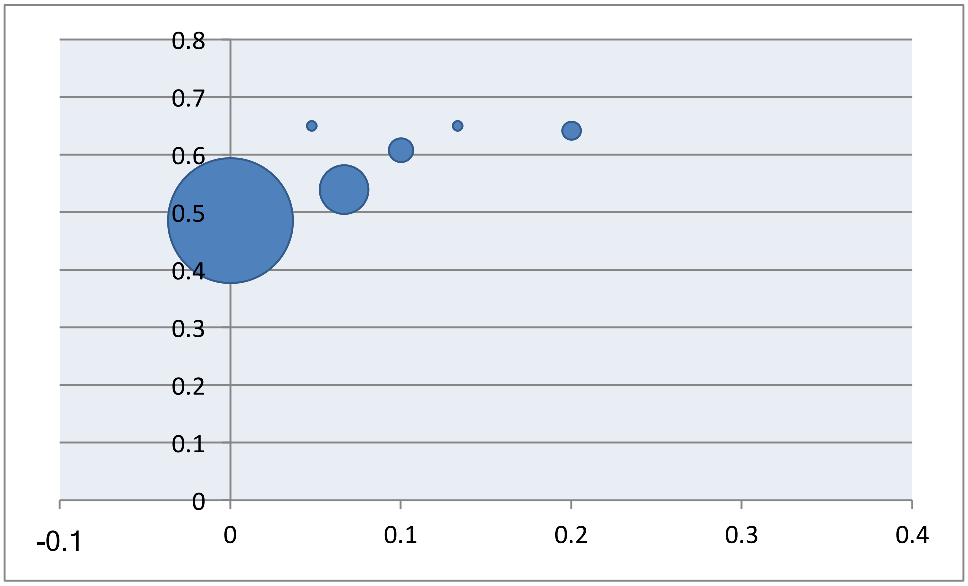


Figure 4. Performance and School Score (2013, 2014-2016)

The Y axis is the performance of the team, X axis is the school score, ranges from 0 (most diverse) to 1 (homogenous). The size of the bubble is proportion to observation number.

2013 (Average School Score=1.70%)



2014 (Average School Score=1.83%)

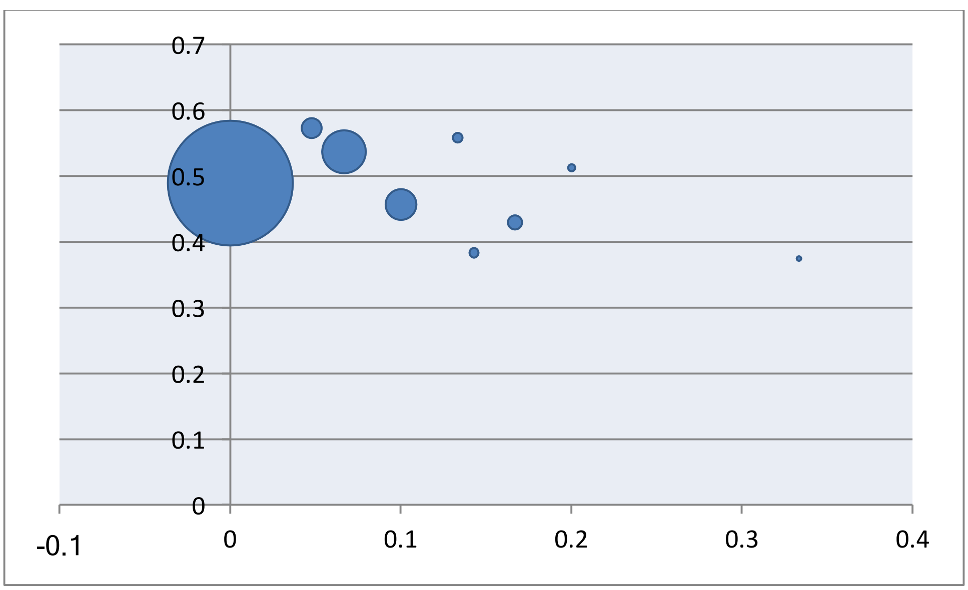
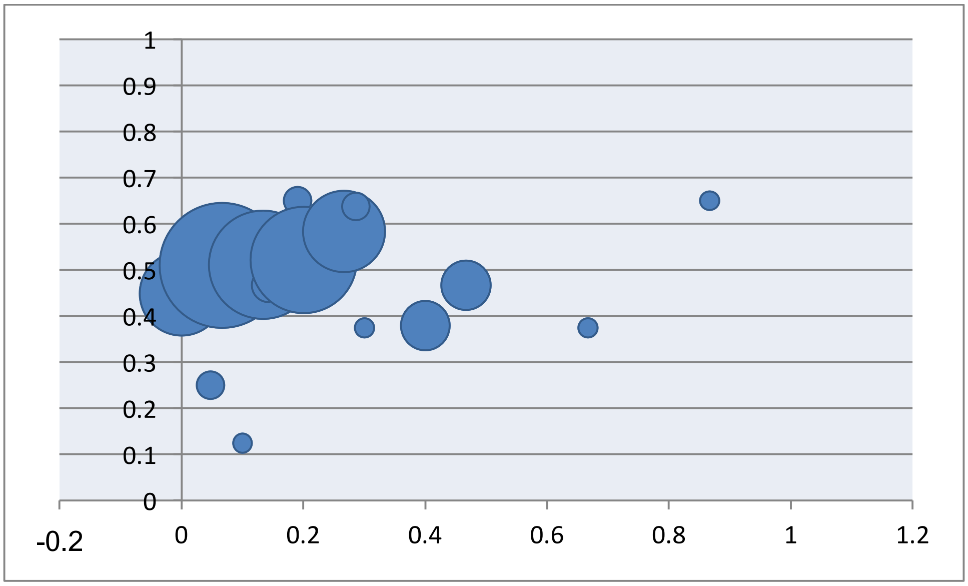


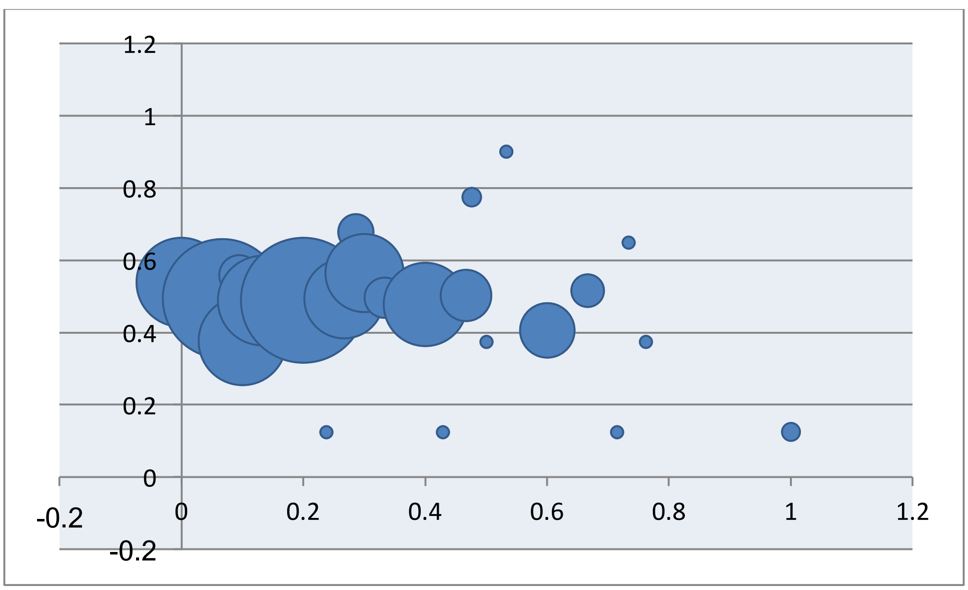
Figure 5. Performance and Industry Score (2013, 2014-2016)

The Y axis is the performance of the team, X axis is the industry score, ranges from 0 (most diverse) to 1 (homogenous). The size of the bubble is proportion to observation number.

2013 (Average Industry Score=16.3%)



2014-2016 (Average Industry Score=21.0%)



Online Appendix Table 1. Home Country/Region of MBA Students

This table reports the home country of HBS MBA students in our sample.

	Country	Freq.	Percent
1	USA	2,394	65.39%
2	India	172	4.70%
3	Canada	125	3.41%
4	China	76	2.08%
5	United Kingdom	59	1.61%
6	Brazil	52	1.42%
7	Australia	46	1.26%
8	France	45	1.23%
9	Germany	45	1.23%
10	Israel	33	0.90%
11	Korea	30	0.82%
12	Japan	28	0.76%
13	Mexico	28	0.76%
14	Turkey	28	0.76%
15	Argentina	27	0.74%
16	Lebanon	25	0.68%
17	Russia	25	0.68%
18	Spain	24	0.66%
19	Nigeria	23	0.63%
20	Chile	19	0.52%
	Total	3,661	100.00%

US	**Obs**	**Percent**
White[15]	1,591	45.46%
Asian American	459	13.11%
African American	199	5.69%
Hispanic American	152	4.34%
International		
European	308	8.8%
South Asian	243	6.94%
East Asian	161	4.6%
Latin American	161	4.6%
Middle Eastern	133	3.8%
African	54	1.54%

[15] We group Canadians (3.41%) and Australians (1.26%) into white Americans.

Online Appendix Table 2. Ethnicity Match by Year

This table reports the probability of matching by dimension by year. “Match” is the probability of being matched to a person with same ethnicity. “Not Match” is the probability of being matched to a person with different ethnicity. “Bias” equals to “Match” minus “Not Match”. “Ratio” equals to “Match” divided by “Not Match”.

Classyear	Match	Not Match	Bias	Ratio
Probability of Matching Conditional on Being a White MBA				
2013	5.65%	5.64%	0.01%	100.14%
2014	6.44%	5.09%	1.35%	126.47%
2015	6.79%	4.99%	1.80%	136.18%
2016	5.40%	3.93%	1.47%	137.43%
2014-2016 Average	6.21%	4.67%	1.54%	133.00%
Probability of Matching Conditional on Being an Asian MBA				
2013	5.80%	5.56%	0.24%	104.26%
2014	7.90%	5.43%	2.47%	145.38%
2015	5.71%	5.67%	0.04%	100.78%
2016	5.84%	4.41%	1.43%	132.42%
2014-2016 Average	6.48%	5.17%	1.31%	125.40%
Probability of Matching Conditional on Being an African American MBA				
2013	5.80%	5.76%	0.04%	100.68%
2014	7.89%	5.49%	2.40%	143.68%
2015	6.90%	5.58%	1.32%	123.68%
2016	4.00%	4.60%	-0.60%	87.00%
2014-2016 Average	6.26%	5.22%	1.04%	119.93%
Probability of Matching Conditional on Being a Hispanic MBA				
2013	6.12%	5.69%	0.44%	107.65%
2014	11.29%	5.40%	5.89%	208.99%
2015	3.70%	5.62%	-1.92%	65.86%
2016	0.00%	4.67%	-4.67%	0.00%
2014-2016 Average	5.00%	5.23%	-0.23%	95.52%
Probability of Matching Conditional on Being an International MBA				
2013	4.07%	5.70%	-1.63%	71.34%
2014	9.23%	5.49%	3.74%	168.08%
2015	8.47%	5.55%	2.92%	152.72%
2016	9.33%	4.35%	4.99%	214.72%
2014-2016 Average	9.01%	5.13%	3.88%	175.72%

Online Appendix Table 3. Match between International Students

This table reports the regression results of the probability of match among international students. Each observation is a student-student pair. The dependent variable real match equals to 1 if the students are teammates. The independent variables are ethnicity characteristics equals to 1 if both students are from the same region.

	2014-2016	2013
	(1)	(2)
VARIABLES	Real_Match	Real_Match
Both European	0.0247***	-0.00892
	(0.00721)	(0.00769)
Both South Asia	0.0286***	-0.0273***
	(0.00849)	(0.00840)
Both East Asia	0.0631***	-0.0401***
	(0.0162)	(0.0114)
Both Latin American	0.0618***	-0.0106
	(0.0192)	(0.0218)
Both Middle East	0.0668***	0.0432
	(0.0178)	(0.0485)
Both African	-0.0474***	-0.0575***
	(0.00148)	(0.00103)
Team Member Count	0.0109***	0.0106***
	(4.85e-05)	(8.86e-05)
2015.ClassYear	-0.00107***	
	(5.88e-05)	
2016.ClassYear	-0.000971***	
	(6.77e-05)	
Constant	-0.0106***	-0.00747***
	(0.000321)	(0.000561)
Observations	254,318	81,368
R-squared	0.002	0.000

Online Appendix Table 4. Probability of All Male/Female/White Teams

This table reports the number of all male/female/white teams by class year. Benchmark is the number of all male/female/white teams if the matching process is random. Each observation is a team.

Class Year	Observed Num of All Male Team	Benchmark	Observed/Benchmark	Obs
2013	0.00	7.43	0.000	150
2014	20.00	6.50	3.077	150
2015	16.00	5.78	2.768	150
2016	24.00	11.62	2.066	180
2014-2016 Average	20.00	7.97	2.511	480
Class Year	**Observed Num of All Female Team**	**Benchmark**	**Observed/Benchmark**	**Obs**
2013	0.00	0.54	0.000	150
2014	7.00	0.64	10.942	150
2015	6.00	0.66	9.048	150
2016	12.00	1.97	6.101	180
2014-2016 Average	8.33	1.09	7.645	480
Class Year	**Observed Num of All White Team**	**Benchmark**	**Observed/Benchmark**	**Obs**
2013	1.00	1.85	0.541	150
2014	2.00	1.95	1.028	150
2015	4.00	1.51	2.653	150
2016	8.00	3.27	2.449	180
2014-2016 Average	4.67	2.24	2.083	480

Online Appendix Table 5. Simulation Result

This table reports simulation results of “ethnicity score” and “school score” under the assumption of random matching within each section. For simplicity, we assume each section is identical, and has 90 students in simulation. The team size is 6. Panel A and Panel B reports the distribution of ethnicity and college of a representative section. Panel C reports the simulation result.

Panel A. Ethnicity	Number	Percent
White	41	45.56%
Asian American	12	13.33%
European	8	8.89%
South Asian	6	6.67%
African American	5	5.56%
East Asian	4	4.44%
Latin American	4	4.44%
Hispanic American	4	4.44%
Middle East	3	3.33%
African	1	1.11%

Panel B. Top College	Number	Percent
Harvard University	7	7.78%
Stanford University	4	4.44%
University of Pennsylvania	4	4.44%
Yale University	3	3.33%
Princeton University	3	3.33%
Duke University	2	2.22%
Massachusetts Institute of Technology	2	2.22%
United States Military Academy	2	2.22%
Dartmouth College	2	2.22%
University of California	2	2.22%
Cornell University	2	2.22%
Georgetown University	2	2.22%
Schools with less than 2 % of MBA population are not listed		

Panel C. Simulation Result	(Iteration=10,000)	
	Mean	SD
Ethnicity Score	0.240	0.175
School Score	0.011	0.029

Online Appendix Table 6. Excess Homophily Score

This table reports regression result of excess homophily on team performance. Independent variable excess homophily is the team homophily score minus benchmark homophily score.

2013 VARIABLES	(7)	(8)	(9) Performance	(10)	(11)	(12)
Excess Ethnicity Score	0.465***				0.465***	0.428***
	(0.142)				(0.151)	(0.150)
Excess Gender Score		0.246			0.0823	0.127
		(0.568)			(0.518)	(0.518)
Excess School Score			0.961*		0.971*	0.827
			(0.559)		(0.535)	(0.593)
Excess Industry Score				0.0932	0.105	0.0918
				(0.151)	(0.153)	(0.161)
Top School Ratio						0.0805
						(0.111)
Start-up Ratio						0.545
						(0.382)
Team Mem Count	0.0495	0.0401	0.0503	0.0440	0.0572	0.0544
	(0.0650)	(0.0711)	(0.0687)	(0.0698)	(0.0624)	(0.0661)
Constant	0.205	0.278	0.199	0.236	0.166	0.136
	(0.395)	(0.440)	(0.418)	(0.425)	(0.387)	(0.416)
Observations	150	150	150	150	150	150
R-squared	0.079	0.003	0.021	0.004	0.101	0.116
2014-2016 VARIABLES	**(1)**	**(2)**	**(3) Performance**	**(4)**	**(5)**	**(6)**
Excess Ethnicity Score	0.172***				0.181***	0.165***
	(0.0534)				(0.0535)	(0.0537)
Excess Gender Score		0.0273			0.0443	0.0649
		(0.0585)			(0.0581)	(0.0586)
Excess School Score			-0.101		-0.199	-0.301
			(0.296)		(0.291)	(0.297)
Excess Industry Score				-0.116	-0.128	-0.118
				(0.0811)	(0.0826)	(0.0828)
Top School Ratio						0.0909
						(0.0559)
Start-up Ratio						0.341**
						(0.134)
2015.ClassYear	0.0437	0.0424	0.0418	0.0379	0.0406	0.0359
	(0.0329)	(0.0332)	(0.0333)	(0.0332)	(0.0327)	(0.0328)
2016.ClassYear	0.125***	0.136***	0.137***	0.140***	0.130***	0.115***
	(0.0392)	(0.0390)	(0.0391)	(0.0394)	(0.0396)	(0.0397)
Team Mem Count	0.0885***	0.0991***	0.0997***	0.105***	0.0946***	0.0850***
	(0.0277)	(0.0274)	(0.0275)	(0.0275)	(0.0276)	(0.0273)
Constant	-0.0876	-0.147	-0.148	-0.176	-0.124	-0.112
	(0.169)	(0.168)	(0.168)	(0.169)	(0.169)	(0.166)
Observations	480	480	480	480	480	480
R-squared	0.048	0.030	0.030	0.034	0.055	0.073