Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Does Social Proximity Enhance Business Partnerships? Theory and Evidence from Ethnicity's Role in U.S. Venture Capital

Deepak Hegde, Justin Tumlinson

MANAGEMENT SCIENCE
Vol. 60, No. 9, September 2014, pp. 2355–2380
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2013.1878

Does Social Proximity Enhance Business Partnerships? Theory and Evidence from Ethnicity's Role in U.S. Venture Capital

Deepak Hegde
Stern School of Business, New York University, New York, New York 10012, dhegde@stern.nyu.edu

Justin Tumlinson
Ifo Institute, University of Munich, 81679 Munich, Germany tumlinson@ifo.de

We develop a formal model to understand the selection and influence effects of social proximity (homophily) between business partners. Consistent with the model's predictions, we find that U.S. venture capitalists (VCs) are more likely to select start-ups with coethnic executives for investment, particularly when the probability of the start-up's success appears low. Ethnic proximity between VCs and the start-ups they invest in is positively related to performance, measured by the probability of the companies' successful exit through acquisitions and initial public offerings (IPOs) and net income after IPO. Two-stage regression estimates suggest that these positive performance outcomes are largely due to *influence*, that is, superior communication and coordination between coethnic VCs and start-up executives *after* the investment. To the extent that VCs expect to work better with coethnic start-ups, they invest in coethnic ventures that are of *lower* observable quality than noncoethnic ventures.

Keywords: venture capital; entrepreneurship; homophily; social capital; social networks
History: Received July 23, 2012; accepted October 13, 2013, by Lee Fleming, entrepreneurship and innovation. Published online in *Articles in Advance* May 21, 2014.

1. Introduction

In 2004, Vinod Khosla, Indian billionaire and co-founder of Sun Microsystems, started Khosla Ventures. By 2011, the Silicon Valley–based venture capital firm's portfolio included U.S. companies founded or cofounded by Ramesh Chandra (MokaFive), Srini Devadas (Verayo), Yogi Goswami (Sunborne), Sandeep Gulati (Zyomed), Siraj Khaliq (WeatherBill), Ramu Krishnan (Ramu Inc.), Ashok Krishnamurthi (Xsigo), Hosain Rahman (Aliph), Anil Rao (Sea Micro), Mulpuri Rao (Soladigm), Bindu Reddy (MyLikes), Mohit Singh (Seeo), and Adya Tripathi (Tula). If we added CEOs' and directors' names, the list of executives of Indian origin in Khosla's portfolio of companies would grow longer still. Khosla Ventures does not advertise a preference for investing in companies started by ethnic Indians, but casual observation suggests that it has one. Is this a costly indulgence of discriminatory preferences, a clever business strategy taking advantage of superior social capital, neither, or both? In this paper we examine how social proximity affects both the choice of business partners as well as subsequent performance.

This study models the interaction of two conceptually distinct mechanisms that shape the performance of socially proximate business partnerships: *selection* and *influence*. Individuals may have better access to, and superior information about, opportunities within their social networks—social proximity may thus facilitate business partner *selection*. After forming a partnership, shared norms and discourse may improve coordination and monitoring among socially close individuals—hence, proximity may positively influence the partnership *after* formation. We formalize these mechanisms and generate testable propositions about the circumstances under which socially proximate agents are likely to partner and succeed. That social proximity should produce superior economic outcomes is neither certain nor obvious. Preferences for interactions with socially close individuals may cause agents to discriminate against superior opportunities outside their social networks and laxly monitor partners within their networks; thus, taste-based selection and influence could undermine the economic success of socially proximate relationships (see, e.g., Gompers et al. 2012).

We test our model's predictions over the social proximity induced by shared ethnicity in the context of the business partnerships formed between venture capitalist (VC) partners (VC "partners" are principals who make and monitor investments) and start-up executives using a sample of almost all U.S. venture-backed deals between 1991 and 2010. We assemble the names of 22,000 U.S.-based VC partners and 85,000 U.S.-based start-up executives from the rosters of 2,687 VCs and 11,235 start-ups they funded and classify each partner and executive, based on their family name (surname) and given name, as belonging to one of 10 distinct ethnic groups. Then, for each investment, we compute a binary measure of coethnicity between the investing VC and funded start-up indicating whether the VC and the company have top-level personnel of the same ethnicity. One may wonder from our example above whether Khosla Venture's investments reflect the preferences of Indian venture capitalists and entrepreneurs for the IT sector or Silicon Valley, rather than ethnic proximity among individuals of the Indian community. To control for these factors, we gather information on investment, VC, and company characteristics, including investment amount, geographic clustering, and VC and company industry specialization.

We first show that Khosla Ventures' investment strategy is not unique. VCs are systematically more likely to invest in a start-up when the VC and company have top-level personnel of the same ethnicity. Ethnic proximity's predictive power is highest for early-stage investments, which have *lower ex ante* success rates than more mature companies. Performance also differs with ethnic proximity. Ordinary least squares (OLS) estimates suggest that both the chance of successful IPO or acquisition and post-IPO performance are higher when the VC and company share an ethnic bond. These superior performance outcomes do not merely represent differences across VCs' investing strategies or attributes—using VC fixed effects regressions we find that even *within* a given VC's portfolio, ethnically closer start-ups perform better. The estimated positive effects of coethnicity are driven by executives belonging to the less common but more distinct ethnic communities in the United States (i.e., individuals not of Anglo-Celtic or West European origin).

These results are based on correlations obtained after controlling for the observable characteristics of VCs and companies but do not distinguish between the effects of ethnicity-based selection of high-quality investments and coethnicity's influence on performance through enhanced coordination between investors and entrepreneurs. We try to isolate the influence effects ("treatment effect" in econometric parlance) of coethnicity by employing three separate strategies: (a) an instrumental variables (IV) approach that accounts for omitted variables, such as unobserved VC and company quality, that affect performance through selection; (b) a method developed by Ackerberg and Botticini (2002), also based on IVs, that isolates the effect of exogenous market characteristics unrelated to the influence effects of coethnicity on performance; and (c) a two-stage Heckman model that corrects for a broader set of factors that affect selection (including unobserved quality) while predicting performance. All three approaches yield estimates of coethnic influence substantially larger than OLS estimates and suggest that coethnicity improves performance through strong postinvestment influence.

Our finding that ethnic proximity facilitates VC-company matching, particularly during early funding rounds when the probability of the start-ups' success is low, taken together with our two-stage estimates implies that VCs select coethnic companies (over noncoethnic ones) even when they appear to be of *lower* observable quality. Although counterintuitive, such behavior aligns with theoretical predictions based on our model of shared discourse systems between coethnic partners. The model suggests that because VCs read the signals from coethnic companies more precisely, and because VCs *anticipate* coethnicity's positive postinvestment influence, lower quality signals from coethnic companies suffice to trigger investment.

We subject our core findings to a battery of robustness tests. First, we check whether reverse causality could be driving our estimated correlations—that is, by VCs appointing coethnic executives to their portfolio companies after they perform well. We obtain the tenure of company executives for a subsample of our data from LinkedIn, the world's largest professional networking site, and find that the performance results strengthen when we limit company executives to those present at the time of investment. Second, the estimated positive effect of coethnicity on performance is stronger when the start-up founder and VC partner who sits on the start-up's executive board (and thus monitors the investments) are of the same ethnicity. Third, the performance benefits of coethnicity are particularly strong for first-time entrepreneurs. Fourth, for a subsample of our data, we show that controlling for previous school ties between VC partners and their portfolio executives, a potential correlate of ethnic closeness, does not qualitatively alter our findings. Fifth, we find that coethnic VCs neither invest more money nor take more time than noncoethnic VCs to achieve successful exits. Hence, coethnicity's positive influence appears to stem from postinvestment coordination efficiencies, not from VCs expending additional resources to ensure the success of coethnic investments.

Our study contributes to the research on social associations in at least three ways. First, sociologists have observed that individuals exhibit "homophily"—a tendency to associate with socially similar others—but have stopped short of investigating the performance implications of homophily (McPherson et al. 2001 provide an excellent survey of research on the topic). We formally derive the performance implications of selection and influence on business partnerships that vary in the strength of their social proximity. Our propositions apply to associations based on attributes other than ethnicity (such as geographic proximity or industry specialization) and to a variety of partnerships including those between employer and employee, mentor and apprentice, and even husband and wife.[1] Second, we build on an approach for identifying ethnic information based on individuals' publicly available names pioneered by Kerr (2008) and Agrawal et al. (2008) to extract a fine-grained classification of ethnic groups in a representative sample of U.S. executives and confirm the viability of this approach for large-sample studies of ethnic origins. Third, a growing body of research explores the influence of social networks on economic transactions conducted either *across* national boundaries (e.g., Gould 1994, Bottazzi et al. 2012) or *within* individual ethnic conclaves (e.g., Kalnins and Chung 2006, Fisman et al. 2012). This research leaves open the possibility that the benefits ascribed to social proximity are not due to proximity per se but rather the parties' specialized knowledge, such as a superior understanding of foreign institutions. We uniquely demonstrate that proximity improves performance, very likely by reducing postselection coordination costs, even *within* a country.

2. Theory

2.1. Model

This section presents a formal model of the core partnership tasks, selection and influence, as a function of social proximity. We build on Morgan and Várdy's (2009) model of hiring under statistical discrimination to include *influence* effects and analyze proximity's effect on partnership performance. To maintain consistency with our empirical context, we label the parties of the model as VC and company engaged in an investment partnership. But since the model abstracts away from any particular activity specific to the VC industry, it is general enough to illuminate the consequences of proximity based on any number of individuals' social attributes such as culture, gender, race, ethnicity, alumni networks, and so on and for different types of partnerships. The key elements of the model are that a party desires a partner with whom to develop a successful relationship by (i) *searching* for a suitable partner, (ii) *screening* potential candidates based on observable signals of quality, (iii) *selecting* a partner, and (iv) *influencing* the relationship after commitment. Social proximity facilitates screening and influence and thus has strategic consequences for search, selection, and ultimately the relationship's performance.

In our context, VCs invest in companies they expect to be successfully sold, either to the public or to another firm. Success is a function of two attributes: (i) the company's unobservable quality $\theta \in \{0, 1\}$, where $\theta = 1$ indicates a high-quality company and occurs with prior probability p, independent of all else, and (ii) the quality of the *postinvestment* relationship between the VC and the company.

VCs and companies reside on an n-dimensional metric space, where location represents relative composition in the space of social associations—the ith coordinate is strength of social affiliation i. In the context of venture capital, this may be the proportion of the VC or company personnel with social affiliation i. The strength of the social tie between the VC and the partner then is measured as the distance (e.g., Euclidean or Mahalanobis) between their respective locations in the space of social associations. VCs evaluate potential deals one at a time by targeting *search* at a specific location z_0 on the metric space. The next potential company to be evaluated is more likely to reside closer to the targeted location than further away. Formally, if $y(z_0, z)$ is the social distance between the location of the search target z_0 and an arbitrary point in the space of social associations z, and $f(z_0, z)$ is the density of discovered companies at location z, then $df(z_0, z)/dy(z_0, z) < 0$.

A VC then observes a signal of the discovered company's quality $\hat{\theta} = \theta + \varepsilon_y$, where y is the observable social distance between company and VC, and noise ε_y is distributed $\mathrm{N}(0, \sigma_y^2)$ such that its variance increases in distance (i.e., $d\sigma_y/dy > 0$). In other words, VCs get more precise signals of proximate companies' unobservable quality, allowing them to *screen* proximate companies better.

Social distance also negatively affects a company's success probability, conditional on quality—it succeeds with probability $\theta G(y)$, such that $G'(y) < 0$.[2] Thus, proximity positively influences success. A VC decides to accept the company or reject it, incurring a fixed search cost and targeting a new search. The game ends when an investment is made.[3]

[1] Incidentally, Bratter and King (2008) provide evidence that interracial marriages are more likely to end in divorce compared to marriages within an ethnic community.

[2] Many activities occur in an investment partnership, which we do not explicitly model but decreasing G sufficiently captures. For example, in a model where VC partners and company executives exert effort with cost increasing in social distance (say, because distant parties are more difficult to communicate with) and both benefit from investment success, it is straightforward to show that in equilibrium, mutual effort and probability of success increase in social proximity, which we do in the theoretical appendix. Since these background activities are unobservable in our empirical setting, we abstract from them for parsimony in the main text.

[3] Of course, the game may be repeated many times, but it is static. Evaluating multiple companies in pursuit of a single investment

We explicitly model the strategic behavior of just one party, in our case, the VC, in relationships. In many partnerships both parties are engaged in symmetric searching, screening, selecting, and influencing activities. Since social distance in bilateral partnerships is reduced for one party if and only if it is reduced for the other, and qualitatively the effects of distance are the same for both parties in all respects, our simplification to model only one side of bilateral partnerships is without loss of generality.[4]

2.2. Example

Before formally analyzing the model, we introduce its mechanisms graphically in Figure 1. Suppose that the VC accepts only companies with probability of success greater than 1/3. Consider first the case where social distance's role is limited to screening: high-quality investments succeed with probability $\bar{G}=2/3$, regardless of social distance.[5] Thus, the VC invests if and only if the posterior probability that the investee is high quality is at least 1/2, or alternatively, weakly greater than the posterior probability that the investee is low quality. From Bayes' rule we can write the required condition as $\Pr\{\hat{\theta}\mid\theta=1,y\}p\geq\Pr\{\hat{\theta}\mid\theta=0,y\}(1-p)$.[6] The left-hand side (LHS) is the density of signals the VC sees from y distant companies that are high quality, scaled by the prior probability p that the company is high quality. The right-hand side (RHS) is the density of signals the VCs sees from y distant companies that are low quality, scaled by the prior probability $(1-p)$ that the company is low quality. The LHS and RHS appear as the dark gray and light gray bell curves, respectively, in Figure 1, where the upper pair represents the scaled signal densities of near companies and the lower pair the scaled signal densities of far companies. The scaled densities intersect where a company is equally likely to be high or low quality.[7] Investments in companies with this signal succeed with probability $\Pr\{\theta=1\mid\hat{\theta},y\}\bar{G}=1/3$. Companies with lower signals (which the VC refuses to invest in) succeed less often, and companies with higher signals succeed more. Notice, though, that this threshold signal is greater for socially distant companies ($\hat{\theta}=0.99$ when near versus $\hat{\theta}=1.6$ when close) because noisier signals cause the VC to weight its priors stronger, which are that companies are typically low quality.

Now, let postinvestment influence also depend on social proximity. Suppose that high-quality companies now succeed with 100% probability if they are socially close to their VC but with only 50% probability if far away (i.e., $G(\text{Near})=1$ and $G(\text{Far})=1/2$). To achieve a 1/3 probability of success, the VC can either invest in a close company that is high quality with a probability 1/3 or a distant company that is high quality with probability 2/3. Under our example parameterization, this is equivalent to observing a close company with a signal $\hat{\theta}=0.68$ or a distant one with a signal $\hat{\theta}=2.29$. Dashed vertical lines denote these new threshold values in Figure 1. Although marginal close and marginal far companies succeed with equal probability, the former now is more likely to be low quality because the VC anticipates superior postinvestment influence and tolerates probabilistically lower quality. This equilibrium effect of superior influence plays a significant role in understanding our empirical findings.

2.3. Analysis

Now we formally derive the intuitions exposed in the above example from the model. Suppose a VC has finished researching a company (i.e., y and $\hat{\theta}$ are observed) and must decide to either invest in the company or reject it in favor of searching for and evaluating another. Let ϕ and Φ, respectively, be the pdf and cdf of the standard normal distribution. From Bayes' rule,

$$\begin{aligned}q_y(\hat{\theta}) &= \Pr\{\theta=1\mid\hat{\theta},y\}\\ &= \phi\left(\frac{\hat{\theta}-1}{\sigma_y}\right)p\Big/\left(\phi\left(\frac{\hat{\theta}-1}{\sigma_y}\right)p+\phi\left(\frac{\hat{\theta}}{\sigma_y}\right)(1-p)\right) \quad (1)\end{aligned}$$

is the posterior probability that a company is high quality. Thus, the probability that an investment in a y socially distant company with signal $\hat{\theta}$ will be successful is $t=q_y(\hat{\theta})G(y)$. Inverting this, the signal

$$\hat{\theta}_y(t)=\frac{1}{2}+\sigma_y^2\ln\left(\frac{1-p}{p}\frac{t}{G(y)-t}\right) \quad (2)$$

indicates that a y ethnically distant company has a t probability of success. Since the value of searching again is always identical ex ante, the cutoff signal above which the VC accepts the company always signifies the same optimal posterior success probability t^*. Formally, $\forall\, y,\ q_y(\hat{\theta}_y^*)G(y)=t^*$, or equivalently, $\forall\, y,\ \hat{\theta}_y^*=\hat{\theta}_y(t^*)$.[8]

does not require multiple periods, though it costs more, e.g., in labor costs, to perform due diligence on each.

[4] Our simplification does not cover partnerships with more than two decision makers because being close to one (potential) member of the partnership does not necessarily imply being close to all other members of the partnership. Search strategies could be more complex in such situations because the closeness to other (potential) partners would need to be considered by all parties.

[5] Formally, for all y, $G(y)=\bar{G}=2/3$.

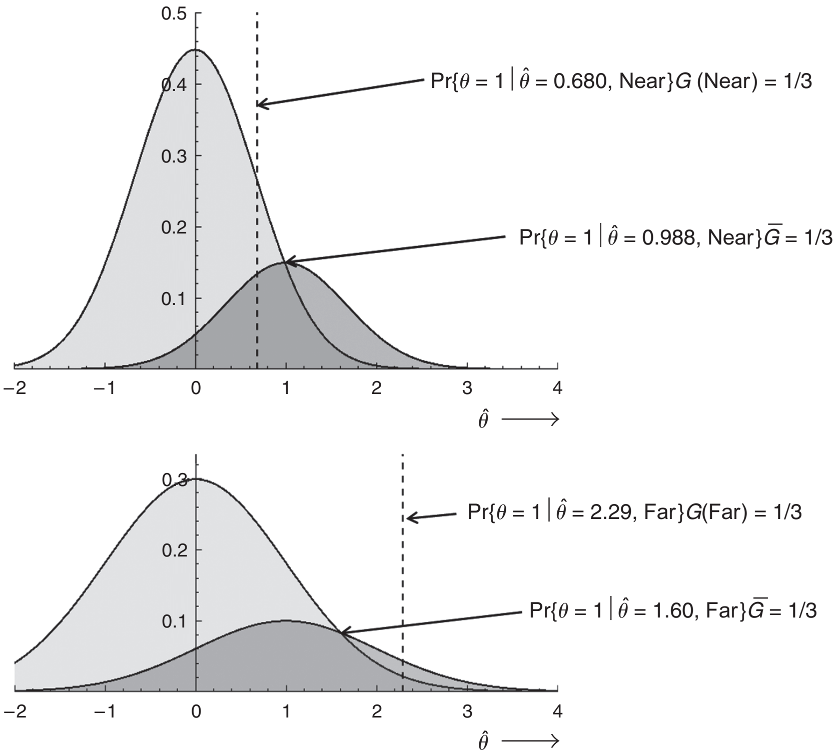
[6] Following Equation (1) in §2.3, this is more completely derived:

$$\Pr\{\theta=1\mid\hat{\theta},y\}=\frac{\phi((\hat{\theta}-1)/\sigma_y)p}{\Pr\{\hat{\theta}\mid y\}}\geq\frac{\phi(\hat{\theta}/\sigma_y)(1-p)}{\Pr\{\hat{\theta}\mid y\}}=\Pr\{\theta=0\mid\hat{\theta},y\}.$$

[7] Formally, $\Pr\{\theta=1\mid\hat{\theta},y\}=1/2=\Pr\{\theta=0\mid\hat{\theta},y\}$.

[8] Explicit calculation of t^* would require additional assumptions on the search costs of VCs, which our data do not inform. Nevertheless,

Figure 1 **Screening and Influence's Role in Selecting Investments**



Notes. This figure depicts the densities of signals observed by a VC, conditional on a start-up being of a particular quality, scaled by the prior probability that a start-up is of that quality. The upper pair of densities occurs when start-ups are socially close to the VCs (i.e., $\sigma_{\text{Near}} = 2/3$), and the lower density pair occurs when start-ups are distant (i.e., $\sigma_{\text{Near}} = 1$). The darker densities are those conditional on the start-up being high quality scaled by the prior probability that start-ups are high quality (i.e., $p = 1/4$), and the lighter densities are those conditional on the start-up being low quality scaled by the prior probability that start-ups are low quality (i.e., $1 - p = 3/4$). Assume that the VC invests in start-ups if and only if the probability of success exceeds 1/3. First, suppose that postinvestment influence (of high-quality start-ups) is $\bar{G} = 2/3$, regardless of social distance. Then the intersection of the dark and light densities determines the threshold signals. Observe that the VC accepts lower signals from close start-ups, but the probability of success *and* the expected company quality at each threshold signal is identical. Now assume that the probability of successful postinvestment influence (of high-quality start-ups) is 100% for close start-ups but only 50% for distant ones. Vertical dashed lines denote the new thresholds—VCs accept relatively even lower signals from near start-ups than when influence was independent of social proximity. But now, although the overall probability of success of start-ups at these respective thresholds is identical, the quality of marginally accepted close start-ups is lower because the VC anticipates smoother postinvestment influence for close companies.

Note that if t exceeds the probability of successful influence activities $G(y)$, an overall probability of success t is simply unattainable. Thus, without loss of generality, we can assume $t^* > G(y)$ because any company not meeting this requirement will automatically be rejected by the VC.

Taking the derivative of (2) with respect to y and evaluating at t^* yields

$$\frac{d}{dy}\hat{\theta}_y(t)\bigg|_{t=t^*} = 2\sigma_y \frac{d\sigma_y}{dy} \ln\left(\frac{1-p}{p}\frac{t^*}{G(y)-t^*}\right) - \frac{\sigma_y^2 G'(y)}{G(y)-t^*}. \tag{3}$$

The first term of (3) is positive if and only if the target posterior probability exceeds the prior probability of success ($pG(y) < t^*$), and the second is always positive.[9] Equation (3) positive implies that the threshold signal to trigger investment decreases in social proximity. Intuitively, the first term denotes the screening advantage VCs have in evaluating proximate firms—the closer the company is, the more reliably a favorable signal indicates high quality (assuming $pG(y) < t^*$). The second term is the influence effect, the advantage that a close company has in exiting successfully, independent of quality. If influence is large, it is optimal for the VC to accept proximate companies of probabilistically lower quality than it would accept among those further away. Thus, we show the following:

Lemma 1. *VCs set lower acceptance criteria for socially proximate companies if the prior probability that a y distant company will succeed ($pG(y)$) is less than t^*, the threshold success probability, regardless of distance.*

we can compute comparative statics on the quality and performance of investments with respect to social proximity without introducing additional assumptions.

[9] The argument of the logarithm exceeds one if and only if $pG(y) < t^*$.

That is, if VCs reject most companies regardless of distance, which we assume and is generally accepted, then VCs accept lower-valued quality signals from close companies.[10] A casual observer might perceive this as taste-based discrimination, but it is not—VCs set the same minimum success probability for companies at all locations. The quality signal denoting this minimum probability is lower for close companies, both because when a close company sends a *high* quality signal, it indicates a high-quality company with greater certainty than when a distant company does so, and because the VC knows it can compensate for low quality, to an extent, with positive influence after investment. This means that a VC has generally observed quality signals from its socially close investments that are lower. So it is reasonable to ask, "How does social distance affect the performance of *actual investments* that the VC makes?" Define the probability that an accepted company of distance y is high quality as:

$$H_y = \Pr\{\theta = 1 \mid \hat{\theta} \geq \hat{\theta}_y^*, y\}$$
$$= \frac{(1-\Phi((\hat{\theta}_y^* - 1)/\sigma_y))p}{(1-\Phi((\hat{\theta}_y^* - 1)/\sigma_y))p + (1-\Phi(\hat{\theta}_y^*/\sigma_y))(1-p)}.$$

Thus, the success rate of accepted companies is $T_y = H_y G(y)$. In the theoretical appendix we prove the following:

Proposition 1. *Socially proximate investments are more likely to succeed* (*i.e.*, $dT_y/dy < 0$).

VCs search their own social circles because close investments are more likely to succeed. Hence, VCs also evaluate disproportionately more companies with whom they have close social associations. Since quality is location independent, Lemma 1 implies that given *any* candidate stream to evaluate, those investments near the VC will be overrepresented in its portfolio. Thus, the following is immediate:

Proposition 2. *VCs are disproportionately more likely to invest in socially proximate companies*.

2.4. Discussion

Since signals of quality are unbiased, an independent auditor's expected signal $\hat{\theta}_A$ of a portfolio company equals the conditional probability that the investment is high quality. Thus, the auditor's *expected* signal varies with the social distance of the portfolio firm to the VC exactly as the probability that the company is high quality, conditional on investment. Formally,

$$\frac{d}{dy}E[\hat{\theta}_A \mid \hat{\theta} \geq \hat{\theta}_y^*] = \frac{d}{dy}\big(E[1+\varepsilon_A \mid \hat{\theta} \geq \hat{\theta}_y^*, \theta = 1]H_y$$
$$+ E[\varepsilon_A \mid \hat{\theta} \geq \hat{\theta}_y^*, \theta = 0](1-H_y)\big)$$
$$= \frac{dH_y}{dy},$$

which follows because noise contained in the signal observed by the auditor ε_A is independent from the noise contained in the signal observed by the VC; thus, although the variance of the auditor's noise varies with the auditor's position, it always has mean zero. Lemma 4 (in the theoretical appendix) calculates dH_y/dy.[11] Lemma 4 shows that the sign of dH_y/dy turns on the sum of two terms: the first corresponds to screening and is negative (i.e., proximity raises quality), whereas the second (i.e., $-\sigma_y^2 G'(y)/(G(y)-t^*)$) corresponds to influence and is positive (i.e., proximity permits lower quality).[12] Thus, if influence were absent (i.e., $G'(y) = 0$), the VC would *screen*-in higher quality, socially close companies—empirically manifesting as a positive selection effect of proximity. However, as the influence effect increases (i.e., $G'(y)$ becomes more negative), proximity increases the probability of selecting high quality companies less. Intuitively, the VC tolerates probabilistically lower quality, but socially close firms, to capitalize on anticipated postinvestment influence. If influence is strong enough, its positive effect overwhelms screening's negative effect, and the VC's close investments are actually lower quality (though still have overall higher probability of success)—empirically manifesting as a negative selection effect of proximity. This does not imply that proximity-based selection benefits are absent. On the contrary, the VC searches to recruit more ethnically proximate companies and screens them more precisely but, anticipating positive postinvestment influence, tolerates lower expected quality to maximize final probability of success. Both screening and influence effects may be arbitrarily strong, but the sign of the empirical selection effect depends on their *relative* strength. Understanding these subtle links between selection and influence helps interpret our empirical findings.

Our model treats selection and influence effects abstractly enough to analyze investment performance in the presence of varied social associations, but how does coethnicity practically convey these advantages in venture capital investing? These two broad effect

[10] (a) "The typical venture organization receives many dozens of business plans for each one it funds" (Gompers and Lerner 2004, p. 7). (b) The threshold falls as social proximity's postinvestment influence strengthens (i.e., $G'(y)$ becomes more negative). In the limit, as social proximity's influence diminishes to 0 (i.e., $G(y)$ approaches a constant), the statement of Lemma 1 becomes "if and only if." To see this, observe that the second term of Equation (3) goes to zero.

[11] The errors of the auditor's signal have zero mean because they are independent of the VC investment decision.

[12] In particular, see Equation (10).

classes could drive performance in at least four distinct ways: (i) VCs and entrepreneurs may search each other and meet at lower cost owing to being part of the same ethnic network—a preinvestment selection effect. (ii) Once in contact, communication advantages, such as shared language or mutual understanding of the significance of qualifications, markets, and opportunities, may reduce asymmetric information and facilitate mutual screening—a pre-investment selection effect. (iii) The communication advantages may continue postinvestment, making monitoring and coordination less costly. Expectations regarding punctuality, work–life balance, employer–employee loyalty, hierarchy, collective versus individual responsibility, and so on, vary with culture. Thus, when unforeseen circumstances arise, coethnic parties may act more compatibly—a postinvestment influence effect. (iv) Misbehavior by either party is more likely to be observed by shared social networks and communicated and punished within the networks. Thus, the shared ethnic community may curtail opportunistic behavior and reduce monitoring costs—a postinvestment influence effect. Although our empirical tests cannot distinguish among the above channels of influence, we attempt to test the model's propositions and isolate the postinvestment influence effects of coethnicity.

3. Empirical Specification and Sample

3.1. The Empirical Specification

Here, we empirically assess the three main predictions of the model stated in Lemma 1 and Propositions 1 and 2. Rather than following the sequence of derivation, we test Proposition 2 and Lemma 1 first, since they pertain to investment selection, and then Proposition 1, which pertains to postinvestment influence. We test Proposition 2 (and Lemma 1) by estimating the probability that a given VC invests in a company ($\Pr\{y_{c,v}=1\}$) as a function of company c's characteristics (denoted by the row vector C), VC v's characteristics (denoted by the row vector V), and VC-company pair characteristics (denoted by the row vector CV). An ideal test of Proposition 2 should estimate the probability of investment as a function of coethnicity for all companies that a VC evaluated and check whether the VCs disproportionately invested in coethnic companies. But we do not have data on the set of companies that VCs evaluated, and so test whether VCs are more likely to invest in coethnic companies, relative to counterfactual opportunities based on the observable characteristics of VCs and companies. For these estimations, implemented both through a conventional multivariate regressions and propensity score matching (PSM) methods, we construct a sample of all actual VC-company pairs (for which $y_{c,v}=1$) and counterfactual VC-company pairs (for which $y_{c,v}=0$). That is, we estimate,

$$\Pr\{y_{c,v}=1\}=\beta_0+\beta_1 C_c+\beta_2 V_v+\beta_3 CV_{c,v}+\varepsilon_{c,v}. \quad (4)$$

We describe the construction of the counterfactuals in §4.

We then test Proposition 1 after specifying binary measures of performance ($z_{c,v}\in\{0,1\}$) of VC-company pairs as a function of variables in C, V, and CV. Each observation is a unique VC-company pair such that the VC invested at least once in the company (i.e., $y_{c,v}=1$). Hence, we estimate,

$$\Pr\{z_{c,v}=1\mid y_{c,v}=1\}=\beta_0+\beta_1 C_c+\beta_2 V_v+\beta_3 CV_{c,v}+\varepsilon_{c,v}. \quad (5)$$

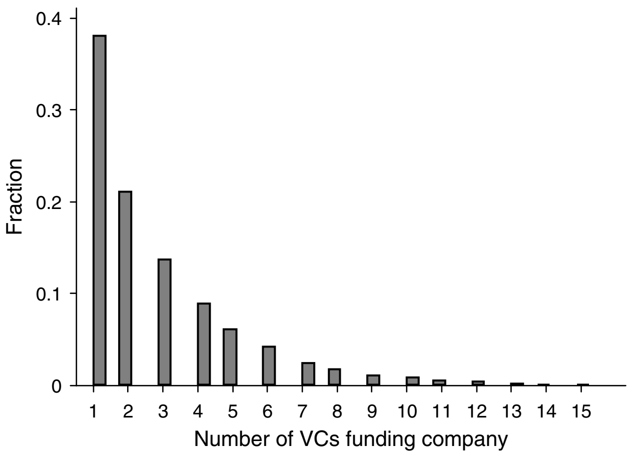
The independent variable of interest in Equations (4) and (5) measures the ethnic proximity of the VC-company pair (an element of CV). We chose the VC-company pair as our unit of analysis rather than the VC partner-company executive pair because investment decisions and contracts are made at the firm (VC/company) level, and our data naturally incorporate information about firm-level attributes that influence the firms' investment decisions (such as industry preference, location, size, and round-level investments). We calculate ethnic proximity between VC-company pairs using information on VC partners' and company executives' ethnic origins. We describe this variable, as well as other elements of C, V, and CV, after describing our estimation sample in detail below.

3.2. The Sample and Variables

We collect data on VCs and their investments from VentureXpert, a proprietary database of Venture Economics owned by Thomson Reuters. Venture Economics assembles data on deals between VCs and their portfolio companies from the quarterly reports of VCs and other institutional investors and supplements these data with information collected from trade publications, company webpages, mailed-out surveys, and telephone contacts with VCs and companies. The coverage of deals in VentureXpert is more comprehensive than in other databases: Gompers and Lerner (2004) conclude that it contains more than 90% of all venture investments, especially for the later years of their study, and Kaplan et al. (2002) report that it covers 85% of all deals.

VentureXpert's information on VCs and companies includes their founding dates, geographic location, industry category, and the names of VC partners and companies' top-level executives. Although VentureXpert covers 290,000 unique deals between 1969 and the present from across the globe, the data on our variables of interest are more complete for the investments of U.S.-based VCs and companies started after 1990. Hence, we restricted our sample to deals covering companies started between 1991 and June 2, 2010, and funded by U.S.-based VCs. This restriction and cleaning the raw data left 2,687 unique U.S.-based VCs

Figure 2 **Histogram of Number of VCs Funding Each Company**



and 11,235 unique U.S.-based companies involved in 73,916 (round-level) deals. The average company in our sample received funding from 2.8 VCs (Figure 2 presents a histogram of the number of VCs funding each company in our sample), and the deals covered 32,017 unique actual VC-company pairs (pairs for which $y_{c,v} = 1$). The following paragraphs describe the construction and sample characteristics of our explanatory and control variables.

3.2.1. Company-Specific Variables. (a) *Ethnic origins of top executives*: VentureXpert lists VC firms' partners and their portfolio companies' top-level executives by given and family names.[13] We assign each executive a most likely ethnicity based on the executive's family name and given name. Origins Info Ltd., a commercial vendor of name-based classification services for ethnically targeted marketing campaigns, provided the assignment. It uses a proprietary database constructed from a variety of sources, such as the *American Dictionary of Family Names* and international telephone directories, to identify the most likely ethnic origin for more than 1,800,000 family names and 700,000 given names. Origins Info's classification assigns an ethnicity to each name based on the family name first and, when family names are inadequate for accurate identification (e.g., for family names like Lee), uses a combination of family name and given name to identify ethnicity (e.g., Seungjun Lee is classified as Korean and Keith Lee as Anglo-Celtic).

Although several studies have validated the accuracy of inferring ethnic origins from names in large samples (see Webber 2007), the approach suffers from several limitations, including that it undercounts the size of ethnic groups whose individuals assume names common among other ethnic groups (e.g., personnel of Jewish origin frequently assume Anglo-Saxon, German, and East European names and are undercounted by our study) and overcounts the size of ethnic groups that provide such assumed names. To the extent that classification errors cause us to miss actual coethnic matches, we expect the errors to make it less likely for us to observe a positive relationship between ethnic proximity and VC-company matching/performance if the relationship actually exists.[14]

The 11,235 U.S. companies in our sample employed a total of 85,168 top-level executives, 13,598 of which were also VC partners, typically listed as nonmanaging board members in the companies. We dropped these executives from the sample of company executives and retained them in the sample of VC partners. Our ethnic classification scheme then assigned each company executive to one of the following 10 most common ethnic groups in the United States: Anglo-Celtic, West European, East European, North European, South European, Chinese, Indian, Japanese, Jewish, and Korean. Executives not belonging to one of the 10 ethnic groups were assigned to a miscellaneous "others" category.[15] Table 1 notes the sub-ethnic groups and nationalities (e.g., English, Irish, and Welsh) that comprise the ethnic groups. Kerr (2008) classifies U.S. inventors using similar ethnic groupings.

Given the completeness of VentureXpert's coverage, our sample distribution of ethnic origins should represent the actual ethnic distribution in U.S. venture-backed companies, subject to the caveats noted above. Table 1 compares the fraction of each ethnicity in the overall U.S. population to the fractions for the executives of U.S.-based start-ups. Top executives of U.S.-based companies are primarily of Anglo-Celtic and West-European origin, more or less comparable to their proportions in the overall U.S. population. Jewish,

[13] VentureXpert does not record the entry and exit of company executives and VC partners, and the list of names we obtained reflects the ethnic composition of companies and VCs when VentureXpert last updated these data. To overcome this limitation, we incorporate information on executives' entry and exit dates from LinkedIn, the professional networking site, for a subset of our data in a robustness check.

[14] To see this, assume that coethnicity is, in fact, associated with positive performance. This means, ceteris paribus, a company executive belonging to ethnicity X is more likely to perform better when she receives investment from a VC partner also belonging to ethnicity X. Classification errors could be of three different types: (i) both the VC and the executive are incorrectly classified as belonging to ethnicity Y, (ii) the executive belonging to X is incorrectly classified as belonging to ethnicity Y, and (iii) the VC belonging to X is incorrectly classified as belonging to ethnicity Y. Type (i) will not affect the estimated average effect of coethnicity, but both types (ii) and (iii) will result in matches that are noncoethnic and a higher probability of superior performance, thus biasing our estimates of the positive effect of coethnicity *downward*.

[15] Table A1 of the online data and methods appendix (available at http://ssrn.com/abstract=1939587) lists the 10 most common surnames for each ethnic group in our sample of U.S.-based company executives.

Table 1 Ethnic Origins of U.S.-Based VC Partners and Executives of Start-up Companies

Ethnic group	1 U.S. overall	2 U.S.-based executives	3 U.S.-based VC partners
Anglo-Celtic[a]	58.45	52.68	50.47
West European[b]	14.3	18.33	18.44
South European[c]	9.76	7.07	6.95
East European[d]	3.32	4.2	4.32
North European[e]	3.2	3.64	3.34
Jewish	0.99	3.62	3.53
Chinese	0.74	1.82	2.96
Indian	0.66	3.53	3.74
Korean	0.37	0.47	1.09
Japanese	0.24	0.5	0.95
Others[f]	7.97	4.14	4.22

Notes. Columns 1–3 display the percentage of individuals belonging to each of the 10 different ethnic origins (and a miscellaneous "others" category) in the following three samples: the U.S. population, the top executives of U.S.-based start-up companies funded by U.S.-based VCs, and U.S.-based VCs. The numbers in column 1 were provided by Origins Info Ltd. based on its records of individuals in the U.S. population; the numbers in column 2 are based on the names of 85,168 top-level executives at 11,235 U.S.-based companies started between 1991 and 2010 and funded by U.S.-based VCs; and the numbers in column 3 are based on the names of 22,110 partners working at 2,687 U.S.-based VCs.

[a]Anglo-Celtic includes individuals with origins in England, Australia, Ireland, Scotland, and Wales.

[b]West European includes individuals with origins in Belgium, Germany, France, the Netherlands, and Switzerland.

[c]South European includes individuals with origins in Greece, Italy, Portugal, and Spain.

[d]East European includes individuals with origins in Albania, the Balkans, Bosnia and Herzegovina, Bulgaria, Croatia, the Czech Republic, Estonia, Hungary, Georgia, Latvia, Poland, Romania, Russia, Serbia, and Ukraine.

[e]North European includes individuals with origins in Denmark, Finland, Iceland, Norway, and Sweden.

[f]Others is a miscellaneous category and includes individuals with origins in Middle Eastern, South American, and South Asian countries not captured by the remaining groups.

Chinese, and Indian individuals are overrepresented as executives relative to their overall populations in the United States.

For each portfolio company c in our sample, a unit vector $\mathbf{e}_c = (e_{c,1}, e_{c,2}, \ldots, e_{c,11})$ indicates its position in 11-dimensional ethnic space (one element each for the 10 ethnic groups plus one for others). Each coordinate indicates the proportion of the company's top executives belonging to the corresponding ethnicity. We calculate the ethnic proximity for each VC-company pair using $\mathbf{e}_c$ and include the vector in our estimations to control for the proportion of different ethnicities within firms and VCs.

(b) *Number of top executives*: Portfolio companies in our sample list 8.55 top-level executives, on average (SD = 5.26; range = 1–56). The executives are most commonly designated Chief Executive Officer, Chief Financial Officer, Founder, President, Director, Board Member, and Vice-President. We use the number of executives belonging to each firm to control for its size and capital requirements, both of which may influence the firms' ethnic preferences and performance.

(c) *Founding year*: An "average year" in our sample from 1991 to 2010 produces 690 start-ups. The surge of start-up companies in 1999 and 2000 (1,218 and 971 start-ups, respectively) reflects the "dotcom boom," and the steep drop in foundings during 2008–2009 (348 and 132 start-ups, respectively) reflects the economic downturn and perhaps truncated coverage in recent years (VentureXpert collects data about a company when a VC reports funding it, typically two to three years after its start date).

(d) *Industry*: VCs invest primarily in the Internet (21.7% of sample companies), computer software (20.7%), medical/health (12.7%), communications (8.1%), and biotechnology (7%) industries.[16] Dummy variables for 18 industries control for unobserved industry-specific features.

3.2.2. VC-Specific Variables. (a) *Ethnic origins of VC partners*: We classified the 22,110 partners of the 2,687 U.S.-based VCs in our sample by ethnic origin as described in 3.2.1. Column 3 of Table 1 reports the fraction of VC partners in our sample by ethnic origin. Most partners are of European heritage (Anglo-Celtic and West European ethnicities together account for nearly 70% of the sample's VC partners). Jewish, Indian, Chinese, Korean, and Japanese individuals are overrepresented as partners relative to their overall U.S. populations.

For each VC v in our sample, we generate an ethnic position vector $\mathbf{e}_v = (e_{v,1}, e_{v,2}, \ldots, e_{v,11})$ to calculate the ethnic proximity of each VC-company pair and control for VCs' ethnic composition.

(b) *Number of partners*: VCs in our sample have 8.2 partners on average (S.D. = 12.23, range = 1–246). The number of VC partners imperfectly proxies for VC size and the depth of its pockets, which may

[16] Table A2 of the online data and methods appendix reports the industry distribution of the sample companies.

influence both the ethnic composition of the VCs and the probability of investing in any given company.

(c) *Founding year*: The VCs in our sample are, on average, older than the companies; 51% were founded before 1991. An "average year" between 1991 and 2010 produces 2.5 new VCs; however, increased VC foundings accompany the surge of start-ups in 1999 and 2000 (nine and six new VCs, respectively). Founding year dummies control for year-specific economic activity that may influence both investments and the ethnic composition of VCs (such as the boom of software start-up companies during the late 1990s that may have increased both VC investments and the fraction of relevant Chinese and Indian personnel).

3.2.3. Company-VC Pair Specific Variables. (a) *Geographic distance*: VCs tend to invest in geographically close companies because collocation, like coethnicity, arguably facilitates superior monitoring and management of investments (Lerner 1995, Sorenson and Stuart 2001). To the extent that ethnic communities tend to cluster in space, geographic proximity may correlate to both ethnic proximity and investment performance (see Agrawal 2008 and Kerr 2008). To control for the geographic clustering of ethnic communities, we measure *geographic distance* between each VC-company pair by converting the headquarters addresses reported by VentureXpert to longitude and latitude via the Google Geocoding API and compute great-circle ("as the crow flies") distances between VCs and companies using the Haversine formula (first published by Sinnott 1984, though long known by navigators).

(b) *Industry distance*: VCs find it easier to make and monitor investments in industries in which they have prior experience (Hellmann 2000). VCs and entrepreneurs belonging to certain ethnicities may also share industry aptitude and experience. If so, shared industry expertise may correlate with ethnic proximity and matching/performance. To control for this, we construct a variable of *industry distance* as the percentage of investments that the VC has made in industries other than the one in which the paired company operates (we use VentureXpert's assignment of each company to one of 18 industries). This measure of *industry distance* ranges from 0, when all of a VC's prior investments were in the matched company's industry, to 1, when the VC has no *other* investments in the company's industry.

(c) *Ethnic distance*: Depending on our empirical context and objective, we use three different measures of ethnic proximity. The first measure allows us to compute and compare coethnicity's effect for each of the 10 major ethnic groups. The measure is a vector of 10 binary variables and, for each ethnic group, indicates whether the VC and the company each have an individual of the ethnicity. For example, a VC-company pair composed of a VC with two partners of Indian origin and one partner of Chinese origin and a company with three executives of Indian origin and one partner of Jewish origin will have *Coethnic Indian* turned on to one, whereas the other nine elements of the pair's coethnicity vector will be set to zero. Of the 32,017 unique VC-company pairs in the sample, 91.2% had at least one Anglo-Celtic employee each, and 56.6% of the dyads shared West European heritage. (Column 1 of Table 2 reports the relative frequency of coethnic VC-company pairs for all ethnic groups in the sample.)

Second, some of our estimations require a more parsimonious measure of coethnicity (e.g., for use in two-stage least squares (2SLS) regressions, which instrument for coethnicity). Hence, we create a binary measure of coethnicity indicating whether or not the paired VC and company both had personnel belonging to any one of the eight ethnic groups with distinct identities in the United States (i.e., the variable is set to one if any VC partner and any company executive of the VC-company pair share the *same ethnicity* other than Anglo-Celtic, West European, or others). According to this measure, 46.6% of the VC-company relationships were based on at least one coethnic partnership. In comparison, a more inclusive measure that indicted any shared ethnicity between VCs and companies (i.e., including VC-company pairs consisting of coethnic individuals of Anglo-Celtic, West European, or other

Table 2 **Ethnic Proximity and Probability of VC-Company Match**

Ethnic group	1 Actual VC-company pairs	2 Counterfactual pairs	3 Difference
Coethnic Anglo-Celtic	0.912	0.857	0.055
Coethnic West European	0.566	0.463	0.103
Coethnic South European	0.235	0.149	0.086
Coethnic East European	0.114	0.077	0.037
Coethnic North European	0.103	0.061	0.042
Coethnic Indian	0.098	0.040	0.058
Coethnic Jewish	0.091	0.052	0.039
Coethnic Chinese	0.041	0.016	0.024
Coethnic Korean	0.007	0.003	0.003
Coethnic Japanese	0.004	0.002	0.002
Coethnic other	0.114	0.067	0.047
Coethnic distinct groups[a]	0.466	0.311	0.155
Coethnic indistinct groups[b]	0.955	0.914	0.041
Coethnic all groups	0.970	0.935	0.035
Mahalanobis ethnic distance	10.35	14.15	−3.79

Notes. This table compares sample means for the different measures of coethnicity for actual VC-company pairs (column 1), counterfactual VC-company pairs (column 2), and the difference between the two (column 3). All differences are statistically significant at 95% confidence levels.

[a]For both actual and counterfactual pairs, *Coethnic distinct groups* = 1 if *any* of (*Coethnic South European*, *Coethnic East European*, *Coethnic North European*, *Coethnic Indian*, *Coethnic Jewish*, *Coethnic Chinese*, *Coethnic Korean*, *Coethnic Japanese*) = 1.

[b]For both actual and counterfactual pairs, *Coethnic indistinct groups* = 1 if *any* of (*Coethnic Anglo-Celtic*, *Coethnic West European*, *Coethnic other*) = 1.

origins) would result in 97% of all sample VC-company pairs being marked as "coethnic" and eliminate the variation (in ethnic proximity) required to identify the effects of coethnicity.

Third, we calculate a continuous measure of *ethnic distance* between each VC-company pair as the Mahalanobis distance between their ethnicity position vectors, $\mathbf{e}_c$ and $\mathbf{e}_v$ described under §§3.2.1.a and 3.2.2.a above. Formally, Mahalanobis distance is $d(\mathbf{e}_v, \mathbf{e}_c) = \sqrt{(\mathbf{e}_v - \mathbf{e}_c)^T S^{-1}(\mathbf{e}_v - \mathbf{e}_c)}$, where vectors $\mathbf{e}_v$ and $\mathbf{e}_c$ represent the ethnic positions of VCs and companies, respectively; S is the covariance matrix; and T the matrix transpose operator. The advantage of the Mahalanobis measure is that unlike our binary measure of coethnicity, it accounts for the statistical prevalence of the different ethnicities in the sample as well as co-occurrence of the different ethnicities in the sample. In all our regressions, we specify the Mahalanobis *ethnic distance*, geographic distance, the number of company executives, and the number of VC partners in logs to soften the effect of outliers.

4. Does Ethnic Proximity Affect VC-Company Matching?

4.1. Proximity and Matching

Our matching analysis requires constructing a sample of VC-company pairs—both actual, for which the investment happened, and counterfactual, for which investment could have happened but did not. Since our sample has 2,687 VCs and 11,235 companies, there are more than 30 million theoretically possible pairs, of which 32,017 are actual and the rest counterfactual. To distill this to a computationally manageable number, we eliminate pairs for which the VC never invests in the company's industry—such matches are unlikely by revealed preference. We also eliminate pairs for which the VC was not active in a one-year window on either side of the (first and last) date on which the company received funding. This retains all actual matches but eliminates nearly 50% of the counterfactual ones, leaving about 15 million counterfactual pairs. We work with random samples drawn from this set of actual and counterfactual pairs due to computational constraints. We draw a 10% random sample (1,300,761 pairs, of which 3,520 were actual matches) and, for each pair, calculate *ethnic distance*, *geographic distance*, *industry distance*, and other company and VC characteristics. We do not observe whether the start-ups in our data approached certain VCs for investment but were turned down, and hence we cannot predict the probability of receiving funding as a function of proximity. Instead, we aim to test whether matched VCs and company executives are more likely to be coethnic than unmatched ones, and our set of counterfactuals, representing a sample of random (but feasible) unrealized matches, serves this objective.

Table 2 reveals that coethnic personnel are, on average, more likely for actual VC-company pairs than counterfactual pairs: the difference in matching likelihood is statistically significant (at $p < 0.05$) for all 10 ethnic groups. Next, we formally investigate the relationship between ethnic proximity and the probability of VC-company match (i.e., $\Pr\{y_{c,v} = 1\}$ in Equation (4)) by estimating multivariate maximum likelihood probit regressions. Table 3 reports the results—probit estimates and corresponding marginal effects of the influence of the explanatory variables on the probability of a VC-company match appear under panel A.

Columns 1 and 2 of panel A confirm that after controlling for geographic distance, industry distance, founding-year effects of VCs and companies, the proportion of different ethnic individuals in VCs and companies, and industry-specific effects, coethnicity is positively related to the probability of a VC-company match for all ethnic groups (except for individuals of Anglo-Celtic origin). The positive effect of coethnicity is statistically significant (at $p < 0.05$) for Chinese, Indian, Jewish, and South European ethnicities (the South European group is more homogeneous than are other European groups and is composed primarily of individuals with origins in Italy and Spain). This finding confirms and extends the result first reported in Bengtsson and Hsu (2013) that Chinese and Indian VCs in the United States disproportionately invest in companies started by members of their own community.

Column 4 shows that the *average* marginal effect of a single coethnic pair on matching for members of distinct ethnic groups (Chinese, Indian, Japanese, Jewish, Korean, East European, North European, and South European) is nearly four times coethnicity's effect for the "indistinct" groups (Anglo-Celtic, West European, and others); in fact, coethnicity's estimated effect for the latter does not statistically differ from zero. The magnitude of the marginal effects may appear small (a single coethnic pair increases the probability that a VC invests in the given company by 0.04 percentage points), but the unconditional probability of a VC-company pair match in our sample is 0.25%, implying that an additional coethnic pair is associated with a 16% higher probability of a match—an economically substantial effect. Columns 5 and 6 confirm the positive effect of ethnic proximity using our Mahalanobis measure of ethnic distance.

One drawback of the above method is that it estimates coethnicity's effect by comparing the characteristics of actual VC-company matches to that of many counterfactual matches. A number of these counterfactual matches may not be comparable to the actual matches along characteristics that affect the probability of matching. Hence, we further refine the set of counterfactual VC-company matches by calculating assignment probabilities, or propensity scores, for VC-company

Table 3 Relationship Between Ethnic Proximity and Probability of VC-Company Match

Panel A: Probit regression results

D.V.: *VC-company match* (0/1)	1 Probit	2 *dy/dx*	3 Probit	4 *dy/dx*	5 Probit	6 *dy/dx*
Coethnic Anglo-Celtic	−0.008	0				
	[0.031]	[0.000]				
Coethnic Chinese	0.126**	0.0008**				
	[0.039]	[0.000]				
Coethnic East European	0.03	0.0002				
	[0.030]	[0.000]				
Coethnic Indian	0.148**	0.0009**				
	[0.030]	[0.000]				
Coethnic Japanese	0.066	0.0004				
	[0.099]	[0.001]				
Coethnic Jewish	0.072*	0.0004*				
	[0.029]	[0.000]				
Coethnic Korean	0.033	0.0002				
	[0.079]	[0.000]				
Coethnic North European	0.048†	0.0003†				
	[0.027]	[0.000]				
Coethnic South European	0.070**	0.0004**				
	[0.023]	[0.000]				
Coethnic West European	0.025	0.0001				
	[0.021]	[0.000]				
Coethnic others	0.017	0.0001				
	[0.025]	[0.000]				
Coethnic distinct groups			0.074**	0.0004**		
			[0.017]	[0.000]		
Coethnic indistinct groups			0.03	0.0001		
			[0.037]	[0.000]		
Log *ethnic distance*					−0.085**	−0.0004**
					[0.014]	[0.000]
Log *geographic distance*	−0.120**	−0.0006**	−0.120**	−0.0006**	−0.120**	−0.0006**
	[0.004]	[0.000]	[0.004]	[0.000]	[0.004]	[0.000]
Industry distance	−0.497**	−0.0025**	−0.501**	−0.0025**	−0.502**	−0.0025**
	[0.042]	[0.000]	[0.042]	[0.000]	[0.042]	[0.000]
Log *no. of company executives*	0.077**	0.0004**	0.092**	0.0005**	0.079**	0.0004**
	[0.013]	[0.000]	[0.011]	[0.000]	[0.012]	[0.000]
Log *no. of VC partners*	0.123**	0.0006**	0.136**	0.0007**	0.127**	0.0006**
	[0.013]	[0.000]	[0.013]	[0.000]	[0.013]	[0.000]
Constant	−1.633		−1.671		−1.235	
Company year fixed effects	Y		Y		Y	
VC year fixed effects	Y		Y		Y	
Industry fixed effects	Y		Y		Y	
% ethnic personnel in VC and company	Y		Y		Y	
Likelihood ratio χ^2	2,936.4		2,760.0		2,764.0	
Prob > χ^2	0		0		0	
Observations	1,300,761		1,300,761		1,300,761	

Panel B: Propensity score matching results

Variable	Sample	Treated	Controls	Difference	S.E.
Coethnic distinct groups	Unmatched	0.468	0.311	0.157**	0.008
	ATT	0.468	0.440	0.027**	0.012
Log *ethnic distance*	Unmatched	1.989	2.212	−0.223**	0.016
	ATT	1.989	2.042	−0.053**	0.021

Notes. Panel A displays probit estimates and marginal effects derived from probit estimates (dy/dx) of the relationship between *ethnic distance* and the probability that a VC invested in the start-up company with which it is paired. A VC-company pair is the unit of analysis in the regressions. The dependent variable is set to one for actual VC-company pairs (i.e., pairs for which the VC invested in the company) and zero for counterfactual VC-company pairs. Robust standard errors clustered at the VC level are shown in square brackets. Panel B presents the average differences in ethnic proximity between actual VC-company pairs and counterfactual VC-company pairs after constructing the counterfactual sample through propensity score matching (caliper matching).

$^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$.

matches. The propensity scores are obtained from a logit regression that predicts the probability of a match based on the set of observable VC characteristics, company characteristics, and VC-company pair characteristics (except ethnic proximity) described in §3.2. We then compare the average ethnic proximity of the actual VC-company pairs to the average ethnic proximity of the counterfactual sample that retains only those pairs with matching probabilities comparable to the actual matches.[17]

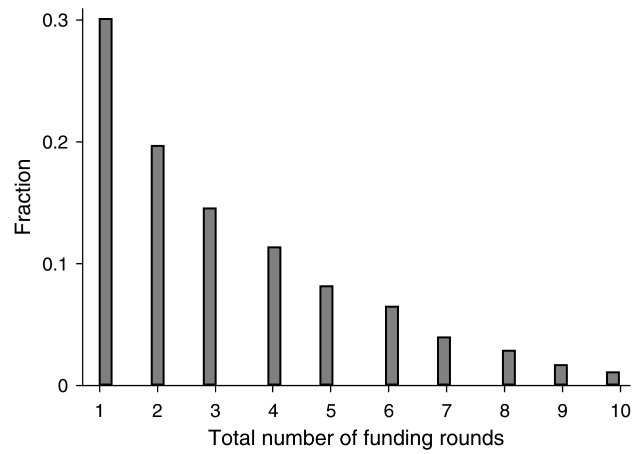
Panel B of Table 3 reports results from this PSM exercise. The estimates suggest that the "average treatment effect on the treated group" (ATT), which is conceptually equivalent to the marginal effect of coethnicity estimated by the probit regressions above, is 0.027 percentage points. This effect is statistically significant, but lower in magnitude than the marginal effect of coethnicity estimated by the probit regression (0.04 percentage points), perhaps because PSM compares against a more plausible set of counterfactuals.[18] The estimated ATT translates to a 11% higher probability of matching for VC-company pairs with coethnic individuals. These results empirically support Proposition 2: VCs match more with (i.e., are more likely to invest in) ethnically close companies.

4.2. Proximity, Matching, and Quality Signals

According to Lemma 1, VCs screen coethnic investments less stringently, both because VCs are surer that the coethnic company they are evaluating is of the indicated quality and because they know that the positive influence effects of coethnicity will compensate for lower quality at the time of investment. Although the econometrician cannot measure the quality signals observed by the VCs when they invested, one can check whether VCs are more likely to invest in coethnic start-ups associated with lower quality signals by using information ex ante generally correlated with start-up success.

Rather than providing all the capital required by start-ups up front, VCs inject capital into their portfolio companies in successive stages, or "rounds." This staged infusion allows VCs to learn about the quality and prospects of start-ups while preserving their option to discontinue funding if the venture appears unlikely to succeed (e.g., Bergemann and Hege 1998, Wang and Zhou 2004). Hence, the average success probability of start-ups at first round funding ($R1$) is lower than the success probability of start-ups that receive second round funding ($R2$), which is lower than the success probability of start-ups that survive to the third round ($R3$), and so on.[19] If VCs are more likely to select coethnic ventures in earlier rounds, then this will provide evidence that VCs tolerate lower quality signals from coethnic start-ups.

Figure 3 **Histogram of Number of Funding Rounds for Each Company**



We construct the actual and counterfactual matches anew to incorporate round-level information. Since in our sample all VCs that fund a given company in round R also fund it in round $R+1$ if the round occurs, we restrict actual pairs to *newly* formed matches and remove counterfactual matches that were actual matches in previous rounds. As before, we only retain plausible matches, based on the companies' industry, the VCs' revealed industry preference, and the relevant window of investment opportunity during which VCs were active. We estimate a separate model for each of the three subsamples representing the first three rounds. Clearly, since survival to round $R+1$ requires survival to round R, and we only consider the new matches in each round, the sample size decreases from round to round. Our computing resources constrained us to work with 30% random samples of the actual and plausible counterfactual matches for each round. The dependent variable for this analysis equals one if the given VC financed the company in the corresponding round and zero if it did not.

Companies in our sample receive 3.3 rounds of funding, on average, and 75% of firms that experience exit events (IPOs, acquisitions, mergers, leveraged buyouts [LBOs], and bankruptcies) do so with five or fewer funding rounds (see Figure 3). So we focus

[17] We experimented with various PSM techniques including nearest neighbor matching, kernel matching, and caliper matching to construct the appropriate control group and confirmed that the results are not sensitive to the technique used. The results reported here are the most conservative ones (i.e., yield the lowest estimates of the effect of coethnicity) and are based on caliper matching, which uses a prespecified tolerance level on the maximum propensity score distance ("caliper") to minimize the risk of bad matches.

[18] We check and confirm that the covariates are identical and balanced across our control and treatment groups. Any differences between the groups are within the acceptable bounds prescribed in Rosenbaum and Rubin (1985). Table A4 of the online data and methods appendix reports the results of our balancing tests.

[19] In our sample, firms that received funding in $R1$, $R2$, $R3$, and $R4$ had IPO probabilities of 7.7%, 9.4%, 11.1%, and 12.2%, respectively.

Table 4 Relationship Between Ethnic Proximity and Probability of VC-Company Match by Funding Round

Panel A: Probit regression results

D.V.: *VC-Company match* (0/1)	1 Probit	2 *dy/dx*	3 Probit	4 *dy/dx*	5 Probit	6 *dy/dx*	7 Probit	8 *dy/dx*
Coethnic distinct groups	0.082**	0.0003**	0.065**	0.0001**	0.050*	0.0001*	0.029	0.0001
	[0.014]	[0.000]	[0.020]	[0.000]	[0.022]	[0.000]	[0.028]	[0.000]
Log *geographic distance*	−0.132**	−0.0004**	−0.095**	−0.0002**	−0.082**	−0.0002**	−0.075**	−0.0001**
	[0.003]	[0.000]	[0.004]	[0.000]	[0.005]	[0.000]	[0.006]	[0.000]
Industry distance	−0.893**	−0.0026**	−0.617**	−0.0013**	−0.509**	−0.0010**	−0.510**	−0.0009**
	[0.033]	[0.000]	[0.048]	[0.000]	[0.059]	[0.000]	[0.068]	[0.000]
Log *no. of company executives*	0.020*	0.0001*	0.043**	0.0001**	0.113**	0.0002**	0.091**	0.0002**
	[0.009]	[0.000]	[0.013]	[0.000]	[0.018]	[0.000]	[0.023]	[0.000]
Log *no. of VC partners*	0.133**	0.0004**	0.112**	0.0002**	0.101**	0.0002**	0.093**	0.0002**
	[0.011]	[0.000]	[0.014]	[0.000]	[0.015]	[0.000]	[0.017]	[0.000]
Constant	−0.891		−1.537		−2.43		−2.291	
Company year fixed effects	Y		Y		Y		Y	
VC year fixed effects	Y		Y		Y		Y	
Industry fixed effects	Y		Y		Y		Y	
% ethnic personnel in VC and company	Y		Y		Y		Y	
Likelihood ratio χ^2	4,217.0		1,810.5		1,080.6		814.7	
Prob > χ^2	0		0		0		0	
Observations	3,001,809		2,151,494		1,479,842		1,000,557	

Panel B: Propensity score matching (PSM) results

Variable: *Coethnic* (*distinct groups*)	Sample	Treated	Controls	Difference	S.E.
Round 1	Unmatched	0.441	0.316	0.124**	0.007
	ATT	0.441	0.411	0.029**	0.010
Round 2	Unmatched	0.499	0.352	0.147**	0.011
	ATT	0.499	0.467	0.032*	0.016
Round 3	Unmatched	0.504	0.375	0.129**	0.014
	ATT	0.504	0.488	0.015	0.020
Round 4	Unmatched	0.499	0.386	0.113**	0.018
	ATT	0.499	0.486	0.013	0.026

Notes. Panel A displays probit estimates and marginal effects derived from probit estimates (*dy/dx*) of the relationship between ethnic distance and the probability that a VC invested in the start-up company with which it is paired separately for the first four rounds of funding. A VC-company pair is the unit of analysis in the regressions. The dependent variable is set to one for actual VC-company pairs (i.e., pairs for which the VC invested in the company) and zero for counterfactual VC-company pairs. Robust standard errors clustered at the VC level are shown in square brackets. Panel B presents the average differences in ethnic proximity between actual VC-company pairs and counterfactual VC-company pairs after constructing the counterfactual sample through propensity score matching (caliper matching).

**$p < 0.01$; *$p < 0.05$.

on the relationship between VC-company matching and ethnic proximity for the first four funding rounds. Table 4 presents the corresponding probit estimates. Because some coethnic groups (e.g., Japanese and Korean) lack sufficient numbers of actual coethnic pairings in each round to precisely estimate their effects, we estimate and report results obtained by using the binary variable that indicates the presence coethnic personnel belonging to any of the eight *distinct* ethnic groups (i.e., coethnic pairs classified as Anglo-Celtic, West European, or others do not set the variable to one).

The estimates in Table 4 suggest that ethnic proximity plays a more significant role in matching VCs to companies during earlier rounds, when VCs face the highest search and screening costs. Panel A shows that an additional coethnic pair is associated with an increase in the probability of matching by 0.03% in the first round (both at $p < 0.01$); for second and third rounds, the effect drops to 0.01% ($p < 0.05$) and does not statistically differ from zero for the fourth round. Although we do not report the estimates for later rounds, we find that the estimated effect of coethnicity for rounds $R5$ and higher was not statistically different from zero. Round-level PSM results also confirm this decay in the estimated ATT with the progression of rounds (see panel B of Table 4). The estimated effects of ethnic proximity follow a similar pattern when measured by the continuous Mahalanobis distance metric. Interestingly, the estimated effects of geographic and industry proximity also follow a similar pattern, consistent with the explanation that search and selection advantages conferred by collocation and cospecialization become

Table 5 Relationship Between Ethnic Proximity and Probability of VC-Company Match by Company Life-Stage

	1	2	3	4	5	6	7	8	9	10
Life-cycle stage:	Seed stage		Early stage		Expansion stage		Late stage		Buyout and acquisition stage	
D.V.: *VC-company match* (0/1)	Probit	*dy/dx*	Probit	*dy/dx*	Probit	*dy/dx*	Probit	*dy/dx*	Probit	*dy/dx*
Coethnic distinct groups	0.111**	0.0003**	0.087**	0.0003**	0.101**	0.0003**	0.074	0.0002	0.019	0.0001
	[0.031]	[0.000]	[0.021]	[0.000]	[0.031]	[0.000]	[0.065]	[0.000]	[0.036]	[0.000]
Log *geographic distance*	−0.146**	−0.0003**	−0.134**	−0.0004**	−0.122**	−0.0004**	−0.142**	−0.0003**	−0.107**	−0.0003**
	[0.007]	[0.000]	[0.004]	[0.000]	[0.005]	[0.000]	[0.012]	[0.000]	[0.007]	[0.000]
Industry distance	−0.967**	−0.0023**	−0.865**	−0.0025**	−0.809**	−0.0026**	−0.844**	−0.0019**	−0.963**	−0.0029**
	[0.069]	[0.000]	[0.042]	[0.000]	[0.057]	[0.000]	[0.120]	[0.000]	[0.058]	[0.000]
Log *no. of company executives*	0.023	0.0001	0.019	0.0001	0.025	0.0001	0.084†	0.0002†	0.008	0
	[0.022]	[0.000]	[0.012]	[0.000]	[0.020]	[0.000]	[0.044]	[0.000]	[0.016]	[0.000]
Log *no. of VC partners*	0.147**	0.0003**	0.123**	0.0004**	0.108**	0.0003**	0.113**	0.0003**	0.199**	0.0006**
	[0.021]	[0.000]	[0.014]	[0.000]	[0.018]	[0.000]	[0.032]	[0.000]	[0.018]	[0.000]
Constant	−1.14		0.109		−1.085		−1.176		−1.201	
Company year fixed effects (19)	Y		Y		Y		Y		Y	
VC year fixed effects (39)	Y		Y		Y		Y		Y	
Industry fixed effects (18)	Y		Y		Y		Y		Y	
% ethnic personnel in VC and company	Y		Y		Y		Y		Y	
Likelihood ratio χ^2	1,573.46		2,754.38		1,250.45		481.49		1,185.67	
Prob > χ^2	0		0		0		0		0	
Observations	620,085		1,338,428		500,846		101,377		422,198	

Notes. This table displays probit estimates and marginal effects derived from probit estimates (*dy/dx*) of the relationship between coethnicity and the probability that a VC invested in the start-up company with which it is paired for companies at different life stages during the first round of funding. A VC-company pair is the unit of analysis in the regressions. The dependent variable is set to one for actual VC-company pairs (i.e., pairs for which the VC invested in the company) and zero for counterfactual VC-company pairs. Robust standard errors clustered at the VC level are shown in square brackets.

$^{**}p < 0.01$; $^{\dagger}p < 0.1$.

less salient as noise about companies' quality decreases.

The probability of start-ups' success also depends on their stage of life. As a start-up matures, ideas become tangible products, business plans translate to verifiable costs and revenues, expansion plans can be better evaluated, and the probability of subsequent failure diminishes. Thus, an alternative test for Lemma 1 suggests that coethnic VCs should be more likely to invest in less mature (i.e., lower ex ante quality) companies. Since the progress of start-ups along their life-cycle correlates highly with the number of investment rounds received, we limit attention to the first time the start-ups receive venture funding—do coethnic VCs invest in less mature companies in $R1$? Of the 10,134 start-ups in our $R1$ sample, 21% were denoted as "seed stage," 41.7% as "early stage," 16.4% as "expansion stage," 3.7% as "late stage," and 17.3% as "buyout and acquisition stage." The estimates in Table 5 confirm that ethnic proximity most significantly predicts VC-start-up matching during the first round of investment for seed stage, early stage, and expansion stage companies (estimated effect of 0.03% at $p < 0.01$ in each case) and has no statistically significant effect for either late stage or buyout and acquisition stage, when the probability of company failure is relatively low.[20]

Finally, the distribution of company age at the time of initial venture investment also indicates that VCs accept lower quality signals from ethnically closer companies. The average start-up company that closes its first funding round with a noncoethnic VC (as before, "coethnic" denotes shared ethnicity among individuals belonging to one of the eight distinct groups) does so 985 days after incorporation compared to 901 days (nearly a full quarter of a year later) for one funded by a coethnic VC. Hence, coethnic investments appear to be associated with lower quality signals, as suggested by Lemma 1.

5. Is Proximity Related to Superior Performance?

5.1. Successful Exits Through IPOs and Acquisitions

Much of the mentoring provided by VC partners to start-ups aims to maximize the likelihood of IPO because

[20] In our sample of firms that received $R1$ funding, those in the buyout and acquisition phase had an IPO probability of 13%, while firms in the earlier stages had IPO probabilities in the 5.7%–8.3% range.

VCs earn the highest average returns through this exit channel (Cochrane 2006, Hochberg et al. 2007). Unlike start-up survival, an alternative measure of success that could be driven by VCs' tastes for keeping ethnically close companies afloat, IPOs require public markets to evaluate company prospects. Therefore, in the absence of investment-level rate-of-return data, IPOs are the clearest available signal of investment success. Although VCs tend to approach the acquisitions market either as a second-best option to going public or when they want to exit a business through "fire sales," previous work suggests acquisitions also generate positive returns for VCs and start-ups (see Gompers and Lerner 2004). Hence, our primary measure of investment success indicates companies' successful exits via IPOs and acquisitions. (The online data and methods appendix shows that our key results reported below hold and, in some cases, strengthen when we restrict our performance measure to indicate exits via IPOs alone.) We also use information on companies' financial performance after IPOs to measure performance for a subset of the companies for which such data are available.

VentureXpert identifies companies that have exited through IPOs, mergers, LBOs, acquisitions, and bankruptcies, but the rest are classified as "Private." Among the companies denoted Private are two types: (i) companies that failed to either go public or be acquired and were eventually written off by the VCs[21] and (ii) companies that were started during the later years of our sample and have not yet had the time to exit or be abandoned. Since many companies designated Private may be defunct and written off by the VCs, we eliminate all Private companies from our analysis that, as of December 31, 2010, had received no funding in more than four years. We chose this threshold because 95% of the companies in our sample that went public or were acquired did so within four years from their last date of financing. We are left with 5,950 unique companies funded by 2,121 VCs and 17,418 observations (unique VC-company pairs) in the estimation sample. We then verify and consider only those events as successful exits for which VentureXpert's indication of IPO and acquisition events were also present in Securities Data Company's (SDC) Global New Issues database and SDC Platinum's M&A database.

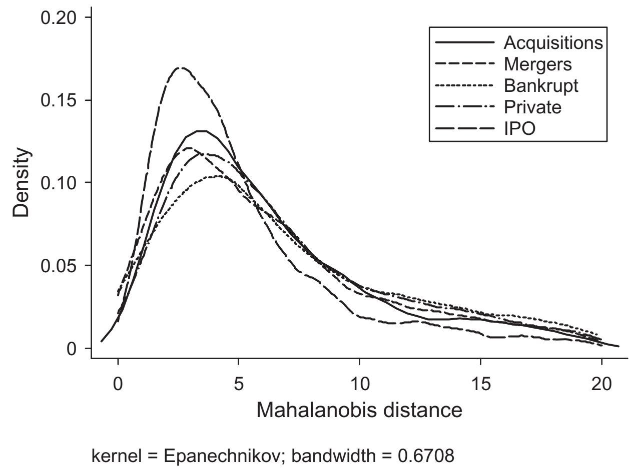
Table 6 shows that 22.2% of the 5,950 companies in our sample exited through IPOs and about the same percentage exited through acquisitions.[22] Thus, overall, 44.5% of the companies are considered successful exits. These companies appear to share a higher proportion of coethnic personnel with their VCs than those that exited through other means or stayed private. Figure 4 also reveals that the distribution of ethnic distances for successful exits, IPOs in particular, is concentrated at lower values.

Table 7 presents probit MLE estimates of the relationship between proximity and successful exits measured by a binary dependent variable equal to one if the company went public or was acquired and zero for all other outcomes. The estimations control for the proportion of VC and company personnel that belong to each of the ethnic groups, industry-specific effects, company and VC founding year effects, size of the business partners, total investment by the VCs in the companies, and geographic and industry distance between VCs and their portfolio companies. As with the matching regressions, we first estimate the effect of coethnicity separately for each of our eleven different ethnic categories. Column 1 shows that shared ethnicity is positively associated with the probability of successful exit for each of our distinct ethnic groups, except for Korean, although coethnicity's positive effect

Table 6 **Company Status and Ethnic Proximity**

Status of companies	% companies	Mean coethnic (distinct groups)	Mean ethnic distance
Went public (IPO)	22.24	0.61	7.50
Acquired or pending acquisition	22.22	0.49	10.30
Private	48.63	0.41	11.95
Merger or LBO	5.88	0.38	12.54
Bankrupt	1.03	0.38	10.25

Figure 4 **Kernel Density Estimates of Ethnic Proximity by Company Status**



[21] Unlike exits through IPOs, mergers or acquisitions, company exits via write-offs or abandonments by VCs are not recorded by VentureXpert. For less than 2% of the companies in the sample, VentureXpert recorded Chapter 7 and Chapter 11 filings, but these numbers do not capture firms that are less officially defunct.

[22] These proportions are similar to the ones presented by Cochrane (2006). In our "raw data," which retains active firms too young to experience exit events, the proportion of firms with IPOs and acquisition events is much lower—12% and 13%, respectively.

Table 7 Relationship Between Ethnic Proximity and Probability of Successful Exit

D.V.: *IPO + Acquired* (0/1)	1 Probit	2 *dy/dx*	3 Probit	4 *dy/dx*	5 OLS	6 OLS	7 OLS
Coethnic Anglo-Celtic	-0.250^{**}	-0.099^{**}					
	[0.054]	[0.021]					
Coethnic Chinese	0.047	0.018					
	[0.063]	[0.025]					
Coethnic East European	0.02	0.008					
	[0.038]	[0.015]					
Coethnic Indian	0.054	0.021					
	[0.044]	[0.017]					
Coethnic Japanese	0.001	0.001					
	[0.147]	[0.058]					
Coethnic Jewish	0.133^{**}	0.053^{**}					
	[0.044]	[0.017]					
Coethnic Korean	-0.295^{**}	-0.111^{**}					
	[0.114]	[0.041]					
Coethnic North European	$0.081^{\dagger}$	$0.032^{\dagger}$					
	[0.043]	[0.017]					
Coethnic South European	0.097^{**}	0.038^{**}					
	[0.032]	[0.013]					
Coethnic West European	$-0.061^{\dagger}$	$-0.024^{\dagger}$					
	[0.033]	[0.013]					
Coethnic others	-0.027	-0.011					
	[0.042]	[0.017]					
Coethnic distinct groups			0.079^{**}	0.031^{**}	0.025^{*}	0.028^{**}	0.037^{**}
			[0.027]	[0.011]	[0.010]	[0.010]	[0.010]
LOG *geographic distance*	-0.006	-0.002	-0.005	-0.002	-0.004^{*}	-0.004^{*}	-0.004^{*}
	[0.005]	[0.002]	[0.005]	[0.002]	[0.002]	[0.002]	[0.002]
Industry distance	-0.311^{**}	-0.122^{**}	-0.310^{**}	-0.122^{**}	-0.070^{*}	-0.080^{*}	-0.079^{*}
	[0.069]	[0.027]	[0.068]	[0.027]	[0.035]	[0.036]	[0.036]
Log *no. of company executives*	0.551^{**}	0.216^{**}	0.522^{**}	0.205^{**}	0.163^{**}	0.175^{**}	0.180^{**}
	[0.024]	[0.010]	[0.022]	[0.009]	[0.007]	[0.007]	[0.007]
Log *no. of VC partners*	0.068^{**}	0.027^{**}	0.046^{**}	0.018^{**}			
	[0.015]	[0.006]	[0.013]	[0.005]			
Log *total funding*	0.115^{**}	0.045^{**}	0.115^{**}	0.045^{**}	0.029^{**}	0.029^{**}	0.029^{**}
	[0.008]	[0.003]	[0.008]	[0.003]	[0.002]	[0.002]	[0.002]
VC board member exists						$-0.020^{\dagger}$	-0.016
						[0.010]	[0.010]
Founder exists							-0.092^{**}
							[0.009]
Constant	-1.102		-1.106		0.324	-0.146	-0.165
Company year fixed effects (19)	Y		Y		Y	Y	Y
VC year fixed effects (39)	Y		Y		Y	Y	Y
VC fixed effects (2007)	N		N		Y	Y	Y
Industry fixed effects (18)	Y		Y		Y	Y	Y
% ethnic personnel in VC and company	Y		Y		Y	N	N
Observations	17,418		17,418		17,418	17,418	17,418
Likelihood ratio χ^2	3,362.19		3,021.79				
Prob $> \chi^2$	0		0				
R-squared					0.289	0.292	0.297

Notes. This table displays estimates of the relationship between ethnic proximity and the probability that the company exits through acquisitions and IPOs. The estimation sample consists of actual VC-company pairs, formed across different rounds of funding (each pair is represented once, regardless of whether the VC funded the company in multiple rounds) and the dependent variable is set to one if the company exited through an IPO or acquisition and zero otherwise. Robust standard errors, clustered at the VC level, are shown in square brackets.

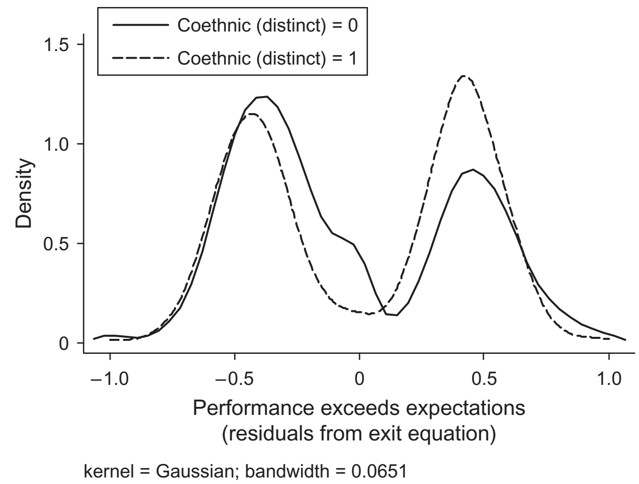
$^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$.

is only statistically significant for the Jewish and South European groups (the lack of statistical significance for the other distinct groups is because of the relatively small number of observations associated with these groups; pooling the information in these coethnic indicator variables into one variable, as our Coethnic distinct groups variable does, increases the number of observations and yields more precise estimates). Interestingly, coethnicity appears to negatively relate to performance for the indistinct groups (Anglo-Celtic and West European). The coefficient on the binary variable that indicates the average effect of shared coethnicity across the eight distinct groups suggests that switching the ethnicity of one VC partner to that of a company executive increases the probability of successful exit by 3.1% (column 4).

Next, we control for the unobserved quality of VC partners by incorporating VC fixed effects (which control not only for VC quality but also other unobserved VC characteristics that may influence their investment performance, such as access to syndicates of co-investors, managerial talent pools, reputation, stage preferences and access to capital). Rather than probit MLE, we estimate VC fixed effects regressions as linear probability models (LPM) for two reasons. First, previous research has shown that slope estimates of nonlinear models with fixed effects, such as probit, can be biased (Heckman 1981); second, our maximum likelihood algorithms for probit fail to converge with the additional 2,121 dummy variables in the VC fixed effects model (this methodological issue, called the "incidental parameters problem," is well documented; see, e.g., Greene 2001). Although the LPM has its limitations, it produces point estimates of the effect of explanatory variables very close to the estimates produced by probit MLE regressions. Column 5 of Table 7 shows that in the model with VC fixed effects, the estimated average effect of a coethnic pair (for distinct ethnic groups) on the probability of successful exit (2.5%) is comparable to the estimated marginal effect of coethnicity without (3.1%).[23] Thus, even within a given VC's portfolio, start-up companies that are ethnically closest to the VCs perform best. To graphically illustrate the effect of coethnicity on performance, we predict the performance of all companies using the full set of controls (i.e., all explanatory variables *except* VC-company coethnicity) and then plot the difference between actual and predicted performances separately for coethnic and noncoethnic VC-company pairs in Figure 5. The density of these residuals for coethnic VC-company pairs is clearly "right shifted," suggesting they perform much better than expected.

[23] In models with VC fixed effects, VC-specific observables such as number of VC partners and percentage of ethnic personnel in VC are subsumed by the VC-specific dummies and drop out of the estimations.

Figure 5 **Kernel Density Estimates of Actual vs. Predicted Performance**



Although any of a start-up's top executives seeking VC funding and several VC partners may be involved in evaluating and selecting investments, interactions between VCs and start-ups after the investment mostly occur between a couple of VC partners and company executives that sit on the company's board. Hence, we recalculate our measure of coethnicity by considering only founders and CEOs of start-up companies and VC partners who sit on their portfolio companies' board of directors and estimate the performance regressions with this refined measure. The companies in our sample list 1.4 top executives on average as either founders or CEOs, and VCs, on average, placed an average of 1.15 partners on companies' boards, conditional on placement. Since only 28% of the VC-company pairs in our estimation sample placed VC partner(s) on the company's board, we calculate the ethnic proximity measure between founders and CEOs of start-up companies and all VC partners for the cases with no VC partners on the company's board and control for the presence of VC partners on the board. Incorporating these changes yields estimates of the effect of coethnicity on the probability of successful exit of 2.8% ($p < 0.01$) for an additional coethnic pairing (see column 6), slightly higher than the estimate (2.5%) obtained using the unrestricted set of company executives and VC partners. If we recalculate the coethnicity measure based on the company personnel that are listed as founders alone (only 22% of our sample pairs were associated with companies that listed at least one founder on their rosters; hence, we also control for the idiosyncratic effect of companies that list the founder), the estimated effect of coethnicity jumps to 3.7% ($p < 0.01$). Since founders could not have been hired at later stages, our estimates of coethnicity's effect on performance are unlikely to be driven by VCs adding coethnic personnel after investment.

We estimate each specification above by restricting successful exits to IPOs alone as well as to IPOs and "good" acquisitions.[24] We find that the corresponding effects of coethnicity are comparable to the ones obtained by including all acquisitions. (The corresponding VC fixed effects estimates are reported in Table A5 of the online data and methods appendix.) The effects of ethnic proximity on performance obtained by using the Mahalanobis measure of proximity are qualitatively similar to the ones obtained by using the binary variable. (Table A6 of the online data and methods appendix tabulates the corresponding results.) As one might expect, the amount of funding received by the companies, VC and company size, geographic proximity, and industry proximity are all positively related to the probability of successful exit.

5.2. Isolating Ethnic Proximity's Influence Effects

The above estimations show that ethnic proximity positively relates to performance, even within a given VC's portfolio. Yet this effect, identified through conditional correlations, could reflect either selection of higher-quality companies by coethnic VCs or positive influence due to coethnic partners' lower coordination costs after investment. Although our matching results suggest that VCs invest in coethnic ventures generally associated with lower quality signals, we do not observe the true quality of the start-ups and cannot immediately conclude that coethnicity has a positive influence on performance.

5.2.1. Instrumental Variables. The main challenge to isolating the influence effects of coethnicity stems from omitted variables, such as the unobserved quality of start-ups, which affect ethnicity-based selection of start-ups and performance. The ideal experiment to identify coethnicity's postinvestment influence would randomly assign start-up companies to different VCs and then measure differences in the performance of coethnic and noncoethnic relationships. However, both our model and first set of empirical findings suggest that VCs do not select start-ups randomly with respect to ethnicity—in fact, we have seen that VCs systematically favor investments in early-stage coethnic start-ups. Alternatively, we could utilize quasi-natural experiments, such as natural disasters or wars, that lead to the migration of ethnic communities into the United States and generate exogenous variation in the availability of coethnic investment opportunities, but no such large-scale "experiments" are available during the period of our study. Thus, we employ instrumental variables to isolate exogenous variation in the probability of coethnic investment; that is, variables that affect the propensity of VCs to invest in coethnic start-ups but do not directly bear upon the postinvestment performance of coethnic relationships.

We propose and implement two IVs: (i) the probability of coethnic investments in a company's market and (ii) state, industry, and year fixed effects at the time of the start-up's founding. Market-level characteristics are natural candidates for IVs in our case because, to an extent, they exogenously determine the availability of coethnic partners and thus the likelihood of coethnic matches. Intuitively, after controlling for geographic proximity and industry specialization of the potential partners, a VC of Indian origin is *more* likely to encounter, and invest in, a coethnic start-up in California (where VCs and entrepreneurs of Indian origin abound) than in New York (where VCs and entrepreneurs are drawn from a broader pool of ethnic backgrounds). However, *conditioned* on encountering a coethnic entrepreneur, an Indian VC is no more likely to enjoy screening advantages in New York than in California. Further, while the average (preinvestment) quality of Indian entrepreneurs may be higher in California than in New York, there is no reason to believe that the average quality of *Coethnic Indian investments* will be higher in California than in New York (unless VCs' preferences to form coethnic matches and the quality thresholds they set to initiate investments differ across states). We can thus examine how, after partialling out the effect of geographic and industry proximity and other observable VC-company characteristics, the variation in the propensity to form coethnic VC-company matches predicted by market elements shapes the performance of coethnic matches. Researchers have previously used similar instruments based on market-level aggregates and fixed effects to identify the treatment effects of variables such as investor experience and geographic proximity on investor performance (e.g., Sørensen 2007, Bottazzi et al. 2008, Tian 2011).

Our first instrument captures variation across markets in VCs' propensity to invest in coethnic companies. We define each "market" as the given company's state, industry, and funding year triplet and calculate the mean ethnic proximity between VCs and companies in each of the 2,875 unique markets in our sample (we exclude the focal VC and company from the calculation of the corresponding market's mean). To facilitate ease of interpretation and to avoid tabulating multiple sets of coefficients, we report OLS rather than probit estimates for the second-stage equations. We start with the baseline OLS estimates with VC fixed effects (reported in column 5 of Table 7 but repeated again in column 1 of Table 8 for easy comparison with the IV estimates). We find that our proposed instrument (average coethnicity of VC-company pairs in the focal firm's market) is strongly related to the probability

[24] We defined "good acquisitions" as those acquisitions for which the transaction value of the acquisition reported by SDC Platinum's Mergers and Acquisition's database exceeded total VC investments in the start-up: 69% of the acquisitions in our data were considered "good" by this measure.

Table 8 Relationship Between Ethnic Proximity and Probability of Successful Exit (Two-Stage Estimates)

D.V.: *IPO + Acquired* (0/1)	1 OLS	2 2SLS	3 2SLS	4 Heckman
Coethnic distinct groups	0.025*	0.121**	0.169**	0.133**
	[0.010]	[0.046]	[0.065]	[0.041]
Log *geographic distance*	−0.004*	−0.004*	−0.003†	−0.178**
	[0.002]	[0.002]	[0.002]	[0.058]
Industry distance	−0.070*	−0.071*	−0.068*	−0.810**
	[0.035]	[0.032]	[0.031]	[0.247]
Log *no. of company executives*	0.163**	0.142**	0.131**	0.241**
	[0.007]	[0.012]	[0.016]	[0.048]
Log *total funding*	0.029**	0.029**	0.029**	0.014**
	[0.002]	[0.003]	[0.003]	[0.003]
Log *no. of VC partners*				0.210**
				[0.067]
Inverse Mills ratio				1.618**
				[0.527]
Constant	0.324	0.791	0.958	−4.031
Company year fixed effects	Y	Y	Y	Y
VC year fixed effects	Y	Y	Y	Y
VC fixed effects	Y	Y	Y	N
Industry fixed effects	Y	Y	Y	Y
% ethnic personnel in VC and company	Y	Y	Y	Y
Observations	17,418	17,418	17,418	3,222
R^2	0.285	0.285	0.279	0.213

Notes. This table displays estimates of the relationship between ethnic proximity and the probability that the company exits through acquisitions and IPOs. The estimation sample consists of actual VC-company pairs, formed across different rounds of funding, and the dependent variable is set to one if the company exited through an IPO or acquisition and zero otherwise. Column 1 presents baseline OLS estimates. Column 2 displays two-stage least squares (2SLS) estimates obtained by using the average of the binary measure of *Coethnic distinct groups* for each focal company's state-industry-funding year as an instrument for *Coethnic distinct groups*. Column 3 displays 2SLS estimates obtained by using fixed effects for the states, industries, and years as well as fixed effects for the interactions of state-industry and industry-funding years as instruments for *Coethnic distinct groups*. Column 4 presents the second stage of the Heckman selection-correction model. The first stage is estimated with the full set of explanatory variables and the instrument used for the estimations in column 2 to satisfy the exclusion restriction; it uses the 10% sample of possible VC-company pairs employed in the matching regressions. Robust standard errors are clustered at the state-industry-funding year level for the estimates in columns 2 and 3 and at the VC level for the estimates in columns 1 and 4 and are shown in square brackets.

$^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$.

of the pair sharing common ethnicity ($\hat{\beta} = 0.522$ at $p < 0.01$; t-statistic $= 49.89$), after controlling for other factors. Column 2 of Table 8 shows that the effect of coethnicity obtained by using this instrument through 2SLS estimation (0.121) is substantially larger than the OLS estimate (0.025) and is statistically significant.

Botazzi et al. (2008) adapt a general IV-based approach proposed by Ackerberg and Botticini (2002) to explain the matching of companies to experienced investors, based on the assumption that the matching of VCs to companies depends on the exogenous market-specific availability of VCs and start-ups. Similarly, the propensity for coethnic matching should differ across markets based on factors unrelated to coethnicity's influence on a given VC-company relationship. Thus, market fixed effects, together with interaction effects among market factors, serve as appropriate instruments to isolate the effects of coethnicity on performance. Our data include companies located in 50 states, in 18 different industries, and funded in the 20 years between 1991 and 2010. We include fixed effects for the states, industries, and years as well as fixed effects for the interactions of state-industry and industry-funding years. This results in 1,275 binary variables ($49 + 17 + 19 + 50 \times 17 + 20 \times 17$) that subsume the effect of quasi-natural experiments, such as changes in visa policy, waves of immigration of certain ethnic communities into the United States, or industry-specific macro trends that arguably influence the probability of coethnic investments, but not their quality, conditional on investment. Column 3 of Table 8 shows that the effect of coethnicity obtained through this approach (0.169) is also substantially larger than the OLS estimate and is statistically significant. Both IV estimations employ standard errors clustered at the state-industry-year levels for the statistical tests. The confidence intervals associated with IV estimates of coethnicity's influence, although estimated with larger standard errors, do not overlap with the confidence intervals around the OLS estimate and are statistically different from the latter (at $p < 0.05$).

We conduct three well-known tests to check the validity of our instruments. First, weak instruments, or instruments that are not sufficiently correlated with the endogenous regressor (the existence of a coethnic bond for a given VC-company pair in our case), will not only fail to correct the biases of OLS estimates but also result in incorrect tests of significance (Bound et al. 1995). The "strength" or relevance of instruments can be checked by testing for the joint significance of the excluded instruments in the first stage; in particular, Stock and Yogo (2005) recommend first-stage F-statistics in the range of 10 to 25 for instrument relevance. We find that the first-stage F-statistics for the excluded instruments in our two IV estimations are 641.8 ($p < 0.01$) and 202.4 ($p < 0.01$), respectively. These values are well above the critical values and pass Stock and Yogo's test for instrument relevance. Second, the Anderson canonical correlation statistics of 1,540.4 ($p < 0.01$) and 755.05 ($p < 0.01$) associated with our two IV regressions also lead us to reject the hypothesis of underidentification, confirming the strong correlation between our excluded instruments and the endogenous regressor. Third, the Durbin-Wu-Hausman test, which involves comparing the coefficients obtained by OLS and IV, rejects the null hypothesis (15.52 at $p < 0.01$ and 11.51 at $p < 0.01$) that the effect of the endogenous regressor is orthogonal to the error term in OLS regressions, thus validating the superiority of estimates obtained through our IV approaches. Still, these IV approaches have their shortcomings: one could argue that coethnic investments in some markets are of systematically higher (preinvestment) quality for reasons that are not adequately captured by our control variables. Subject to this caveat, our IV tests suggest that the influence of unobserved variables, such as quality, on the performance of coethnic partnerships is negative, further supporting the possibility that VCs select coethnic firms of lower (preinvestment) quality than noncoethnic firms.

5.2.2. Heckman Selection Correction. We next follow the approach proposed in Heckman (1979), which explicitly estimates a first-stage selection equation that predicts the sorting and matching of VCs and companies and then incorporates this information in an outcome equation that estimates the treatment effect of coethnicity. To implement Heckman's two-stage model, we return to the sample of possible VC-company matches used to estimate the effect of coethnicity on the probability of matching. As before, after eliminating nonplausible, counterfactual matches as described before, we estimate a first-stage equation with the full set of company, VC, and pair characteristics (explained in §4.1). We also add an additional variable to the matching equation that will not be part of the second-stage outcome equation: the percentage of coethnic matches in each company's founding year-state-industry. This variable relates closely to the matching of VC-company pairs for the reasons explained above, does not directly affect the influence of coethnicity on performance for a given VC-company pair, and thus imposes the exclusion restriction. We then use the parameter estimates obtained by this matching equation to compute the inverse Mills ratio (IMR) for each observation. Finally, using the "selected" sample—observations for which VCs invested in the company—we estimate the outcome equation with our usual set of control variables and the IMR as an additional explanatory variable to correct for selection bias.[25] Column 4 of Table 8 reports the corresponding estimates of the performance equation. The estimated effect of coethnicity (0.13), which we interpret as its influence effect (or "treatment effect"), is again significantly larger than the OLS estimates.

Table A7 of the online data and methods appendix confirms that the strong positive influence effect of coethnicity holds when we restrict the definition of success to exits through IPOs alone. Moreover, the strong positive postinvestment effects of coethnicity persist even after successful exit. Table 9 shows that companies that are ethnically closer to their VCs continue to flourish even after IPO: in the model with VC fixed effects, an additional coethnic pair is associated with, on average, a \$0.1 million higher market capitalization and \$0.009 million higher net income one year after IPO for the start-ups. Thus, we find no evidence that ethnically close VCs and companies "hoodwink" public markets in their IPOs.

These tests collectively confirm coethnicity's strong positive influence on performance. Also, although we have not directly established the magnitude (or direction) of coethnicity's selection effect, the fact that OLS estimates, which include the effects of selection, are substantially lower than the IV estimates that control for unobserved quality suggests that VCs tend to invest in relatively lower-quality coethnic companies. This selection of start-ups associated with lower quality signals is consistent with our finding that coethnic partnerships are particularly likely during early funding rounds and when companies are young. These results align with the intuition exposed by our model: VCs search to recruit more ethnically proximate companies and screen them more precisely but in anticipating positive postinvestment influence tolerate lower expected quality to maximize final probability of success.

5.3. Alternative Explanations and Robustness Checks

(i) We find a strong positive relationship between VCs' investments in coethnic ventures and investment

[25] Our Heckman correction model does not include VC fixed effects both because they do not meaningfully belong in the first stage that predicts matching of VCs to start-ups and because the small number of actual matches available to us in the second stage (since they are obtained from the 10% random sample of possible VC-company pairs used in the first stage) does not permit the estimation of VC-specific intercepts.

Table 9 Relationship Between Ethnic Proximity and Post-IPO Performance

	1	2	3	4
D.V.:	*Market capitalization*		*Net income*	
Coethnic distinct groups	0.091*	0.111*	0.005*	0.009†
	[0.041]	[0.055]	[0.002]	[0.005]
Log *geographic distance*	−0.012†	−0.011	−0.272	−0.897
	[0.006]	[0.010]	[0.389]	[0.812]
Industry distance	−0.302**	−0.370†	9.174	0.658
	[0.093]	[0.191]	[6.723]	[20.550]
Log *no. of company executives*	0.350**	0.311**	2.262	0.204
	[0.036]	[0.048]	[2.066]	[3.846]
Log *total funding*	0.113**	0.147**	−1.758*	−2.498†
	[0.019]	[0.029]	[0.800]	[1.384]
Log *no. of VC partners*	0.036†		2.201†	
	[0.019]		[1.229]	
Constant	1.604	4.154	132.943	−15.687
Company year fixed effects	Y	Y	Y	Y
VC year fixed effects	Y	Y	Y	Y
VC fixed effects	N	Y	N	Y
Industry fixed effects	Y	Y	Y	Y
% ethnic personnel in VC and company	Y	Y	Y	Y
Observations	2,943	2,943	1,316	1,316
R^2	0.326	0.577	0.257	0.553

Notes. This table displays ordinary least squares estimates of the relationship between ethnic proximity and post-IPO performance. The estimation sample consists of 2,943 actual VC-company pairs for companies with data on market capitalization (dependent variable for the estimations in columns 1 and 2 expressed in million dollars) and 1,316 actual VC-company pairs for companies with data on net income one year after IPO (dependent variable for the estimations in columns 3 and 4 expressed in million dollars). Robust standard errors, clustered at the VC level, are shown in square brackets.

$^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$.

success and interpret this as evidence for the positive influence of coethnic partners. But one might ask whether reverse causality drives our estimates: when a company performs well, VC partners replace its top executives with their ethnic brethren. Interviews with VCs and entrepreneurs, however, suggest just the opposite—neither party wishes to alter a successful partnership. Thus, such replacements in thriving firms are rare. Our estimates obtained after limiting company personnel to founders, who are unlikely to be VC chosen replacements, should also mitigate this concern. Still, we examined this issue directly by assembling data on the entry and exit dates of the company executives in our sample from LinkedIn, the world's largest professional networking database. We were able to identify and gather data for 5,272 (6.1%) company executives in our sample that had employment records on LinkedIn. We then recalculated our ethnic distance measure for each VC-company pair using only those executives that were actively employed by the start-up company when the company was first funded by the corresponding VC. This reduces the number of VC-company pairs for which we can compute *ethnic distance* to 1,306, or 7.4% of the full list of actual VC-company pairs. The estimated effects of ethnic proximity obtained by fitting the successful exit-probability regressions to this restricted sample are 2 to 2.5 times larger than those obtained from the corresponding full-sample OLS/probit estimates. (Columns 1 and 2 of Table A8 in the online data and methods appendix display the corresponding estimates.)[26]

(ii) Our analysis estimates the effect of ethnic proximity after controlling for the most salient VC and company characteristics known to affect performance. But our coethnicity measure may pick up the effect of other ethnicity-related social associations, including school ties between VC partners and company executives. (For example, Bengtsson and Hsu 2013 show that VC partners from elite U.S. universities tend to fund ventures founded by executives with degrees from elite U.S. universities; Rider 2012 finds that social associations, including school ties, affect VCs' partnership decisions.) We investigated the effect of common school ties by assembling data from LinkedIn on the educational institutions attended by VC partners and company executives . We coded a binary "school ties" variable indicating whether one of the VC partners attended the same institution as any of its portfolio company's executives (we could construct this variable for 31% of the VC-company pairs in our sample after identifying educational institution affiliations for about 6% of company executives and VC partners). We then reestimated our performance regressions including the school ties variable. Columns 3 and 4 of Table A8 of the online data and methods appendix show that although school ties have a strong positive relationship on the probability of successful exit, they do not qualitatively alter the estimated effects of coethnicity.[27]

(iii) We have argued that the observed positive relationship between proximity and investment performance stems from coethnicity's strong influence effects; that is, coethnic VC partners and company executives work together better because of reduced coordination and monitoring costs. One might argue still that the positive relationship of coethnicity could be driven by VCs allocating more time and resources to coethnic companies. Table A9 of the online data

[26] We cannot estimate VC fixed effects regressions for this subsample because of the small number of observations relative to the number of VCs.

[27] Column 3 of Table A8 suggests that the effect of coethnicity on successful exits (IPOs and acquisitions) is not statistically significant on the inclusion of the school ties variable, but this is an artifact of the subsample for which LinkedIn data on school ties are available rather than due to correlation between school ties and coethnicity. We estimate the identical regression by *omitting* the ties variable using the LinkedIn subsample and find that the estimated coefficient on coethnicity (0.018) is also not statistically significant in this subsample. However, the effect of coethnicity on IPOs remains robust and significant (column 4 of Table A8).

and methods appendix shows that companies that successfully exit when paired with coethnic VCs (a) do not require more time to exit (measured from the first funding round); (b) do not go through more funding rounds; and (c) do not receive more funding. These findings are inconsistent with the argument that VCs inefficiently subsidize their coethnic investments to inflate their probability of success.

(iv) Serial entrepreneurs have prior histories of company founding or success, give off more precise signals of quality, and may need less communication and monitoring from coethnic VCs after investment. Thus, start-ups with serial entrepreneurs are comparable to more mature companies, and we expect coethnicity to play less of a role in the presence of such experienced entrepreneurs. We identify all company executives that appeared on the rosters of three or more companies in our data set as serial entrepreneurs: 13% of our sample companies listed such individuals most likely to be serial entrepreneurs. As expected, excluding the companies associated with these individuals yielded estimates of coethnicity *higher* than the ones obtained from the full sample. (Table A10 of the online data and methods appendix presents the corresponding estimates; we also find that companies associated with the serial entrepreneurs in our sample are nearly twice as likely to successfully exit, ceteris paribus.)

5.4. Effect of Ethnic Proximity on VCs' Payoffs

We find that ethnic proximity of VCs and entrepreneurs is associated with a higher probability of the portfolio investment going public or being acquired. How much is this increased likelihood of IPO or acquisition worth to VCs? We compute the positive impact of an increase in IPO or acquisition probability on the ex ante expected rate of return for an investment as the derivative of expected rate of return with respect to IPO or acquisition probability. First, condition the expected rate of return (r) on the IPO or acquisition event (abbreviated *IPO* below):

$$E[r] = E[r \mid IPO]p + \mathbb{E}[r \mid \text{no } IPO](1-p), \tag{6}$$

where p is the probability of an IPO or acquisition. Then the derivative with respect to p is just the difference between the expected rates of return when an IPO or acquisition occurs and when it does not:

$$\frac{d\mathbb{E}[r]}{dp} = \mathbb{E}[r \mid IPO] - \mathbb{E}[r \mid \text{no } IPO]. \tag{7}$$

Since data on rates of return for individual investments that do not exit via IPO or acquisition are not generally available, we isolate $\mathbb{E}[r \mid \text{no } IPO]$ from (6) and substitute it into (7):

$$\frac{d\mathbb{E}[r]}{dp} = \mathbb{E}[r \mid IPO] - \frac{\mathbb{E}[r] - \mathbb{E}[r \mid IPO]p}{1-p}.$$

Simplifying further,

$$\frac{d\mathbb{E}[r]}{dp} = \frac{\mathbb{E}[r \mid IPO] - \mathbb{E}[r]}{1-p}. \tag{8}$$

Estimates for the three parameters on the RHS of the above can be found in Cochrane (2005). Cochrane estimates mean returns of 698% on VCs' investments that exit in IPOs or acquisitions. Accounting for the selection that occurs prior to a successful exit, Cochrane estimates overall ex ante expected returns to VC investments ($\mathbb{E}[r]$) of 59%. In Cochrane's sample, 41.9% (p) of firms exit via IPO or are acquired (not including an additional 3.7% registered for IPO), comparable to the 44.46% of firms that exit through IPOs or acquisitions in our sample. Substituting these values into (8) yields $d\mathbb{E}[r]/dp = 11.00$. This implies that our conservatively observed increase in the probability of successful exit of 2.5% (column 5 of Table 7) associated with an additional executive who shares ethnicity with a VC partner increases the expected rate of return around 27.5% at the time of investment. These internal rate of return estimates show that the economic returns of coethnic partnerships are substantial, but they should be interpreted cautiously—they rely on Cochrane's finding that VCs, on average, enjoy 698% returns from successful exit events.

6. Conclusion

Our formal model highlights the subtle interaction between the selection and influence effects of social associations in business partnerships. It can be applied to many settings where the association between potential partners can be described with a distance metric. The model proposes that if proximity improves (selection relevant) information and most potential candidates are unsuitable, then increased confidence in their evaluation will cause evaluators to set lower acceptance thresholds over observable quality signals for nearby candidates. If proximity also improves performance after the partnership's formation, then anticipating this, evaluators will drop thresholds for close opportunities further, even to the point that close candidates of lower quality will be accepted. But this is not taste based discrimination—for these close relationships will perform better on average than distant ones. Thus, agents will target their searches for potential partners nearby and partner disproportionately with social neighbors.

Our empirical analysis confirms the model's predictions. We show that conditional on investment, ethnic proximity between VCs and company executives is positively related to the probability that the venture exits in an IPO or acquisition and to post-IPO market capitalization and net income. We also show that VCs are more likely to select ventures led by coethnic executives for investment, and the effect of proximity on investment selection is particularly salient for early-stage start-ups. Thus, our findings suggest that in the VC industry, favoritism toward one's ethnic brethren brings superior economic payoffs. According to the National Venture Capital Association (2009, p. 2), "In 2008, [U.S.] venture capital-backed companies employed more than 12 million people and generated nearly \$3 trillion in revenue." If the ethnicity of a single executive can substantially affect the probability of investment from a particular VC, of growing to sale on public markets and post-IPO income, as we have found, we can conclude that individuals' social associations have profound economic consequences.

In our study, ethnic proximity proxies for a complex web of social ties that include linguistic, religious, and many other associations that bind together members of the same ethnic group. Individuals may choose to tap into certain associations borne out of a common ethnicity and not others. In teasing apart the effects of shared location, industry preferences, and educational background from less-distinct aspects of ethnic proximity that plausibly affect investments, we have only taken a first step in identifying the true effects of ethnic proximity and the channels through which they operate.

Acknowledgments

For helpful suggestions, the authors thank Ashish Arora, Lauren Cohen, Annamaria Conti, Gary Dushnitsky, Gabriel Felbermayr, Ulrich Hege, Thomas Hellman, Johannes Koenen, Tobias Kretschmer, John Morgan, David Mowery, Manju Puri, Chris Rider, Melissa Schilling, Olav Sorenson, Felix Vardy, and Rosemarie Ziedonis, as well as seminar participants at New York University, the University of California, Berkeley, Duke University, Rutgers University, the University of Munich, the National Bureau of Economics Research 2011 Winter Meeting, the 4th HEC Workshop on Entrepreneurship, the Indian Institute of Management (Ahmedabad and Bangalore), and the Strategy Research Initiative/*Administrative Science Quarterly* 2011 Professional Development Workshop. The authors thank Jörg Claussen for his assembly of the LinkedIn data set. The research benefitted from the helpful suggestions of three anonymous referees and the associate editor. This research was supported by the Kauffman Foundation [Grant 20110112] and the Institute of Business and Economic Research at the University of California, Berkeley.

Theoretical Appendix

Definition 1. Let $R(x) = (1-\Phi(x))/\phi(x)$ denote the Mill's ratio.

Lemma 2. *The function $xR(x)$ strictly increases for all x.*

Proof. Case $x > 0$:

$$\frac{d}{dx}(xR(x)) = R(x) + xR'(x) > 0$$

$$\Leftrightarrow \frac{1}{R(x)} > -x\frac{R'(x)}{R(x)^2} = x\frac{d}{dx}\left(\frac{1}{R(x)}\right). \tag{9}$$

From Gordon (1941), $R(x) < 1/x$. From Sampford (1953), $0 < (1/R(x))' < 1$. Thus,

$$\frac{1}{R(x)} > x > x\frac{d}{dx}\left(\frac{1}{R(x)}\right),$$

and the inequalities in (9) hold.

Case $x \le 0$: Clearly, $R'(x) < 0$, for all $x \le 0$, as the denominator $(\phi(x))$ is increasing and the numerator $(1-\Phi(x))$ is decreasing. Thus,

$$\frac{d}{dx}(xR(x)) = R(x) + xR'(x) > 0$$

because $R(x) > 0$, $x \le 0$, and $R'(x) < 0$. □

Corollary 1. *The function $(x-1)R(x)$ strictly increases for all x.*

Proof. Taking the derivative,

$$\frac{d}{dx}((x-1)R(x)) = \frac{d}{dx}(xR(x)) - R'(x) > 0$$

because $(xR(x))' > 0$ from Lemma 2, and $R'(x) < 0$ is well known.

Lemma 3. *For all $z > 0$,*

$$\frac{1}{R(x)} < \frac{1}{z} + \frac{1}{R(x-1/z)}.$$

Proof. Observe that

$$\begin{aligned}\frac{d}{dz}\left(\frac{1}{z} + \frac{1}{R(x-1/z)}\right) &= \frac{1}{z^2}\left(-1 - \frac{R'(x-1/z)}{R(x-1/z)^2}\right)\\ &= \frac{1}{z^2}\left(-1 + \frac{d}{dw}\left(\frac{1}{R(w)}\right)\right) < 0,\end{aligned}$$

where $w = x - 1/z$. From Sampford (1953), $0 < (1/R(w))' < 1$ such that the inequality holds when $w > 0$. Furthermore, since the hazard rate of the normal distribution is well known to increase (i.e., $(1/R(w))' > 0$, $\forall\, w$), the inequality must also hold for $w \le 0$. Since the form of the lemma clearly holds with equality as $z \to \infty$, the RHS of the lemma approaches the LHS from above, everywhere. □

For notational purposes, define $x_1 = (\hat{\theta}_y^* - 1)/\sigma_y$ and $x_0 = \hat{\theta}_y^*/\sigma_y$.

Lemma 4. *The probability that an investment is high quality changes with distance according to*

$$\frac{dH_y}{dy} = \left(\frac{x_1R(x_1) - x_0R(x_0)}{R(x_1) - R(x_0)}\frac{d\sigma_y}{dy} - \frac{\sigma_y^2 G'(y)}{G(y) - t^*}\right)K, \tag{10}$$

where

$$K = \frac{p(1-p)\phi(x_1)\phi(x_0)(R(x_1) - R(x_0))}{\sigma_y((1-\Phi(x_1))p + (1-\Phi(x_0))(1-p))^2} > 0.$$

Proof. From the quotient rule,

$$\begin{aligned}\frac{dH_y}{dy} &= \frac{-\phi(x_1)p(dx_1/dy)}{(1-\Phi(x_1))p + (1-\Phi(x_0))(1-p)}\\ &\quad - H_y\frac{-\phi(x_1)p(dx_1/dy) - \phi(x_0)(1-p)(dx_0/dy)}{(1-\Phi(x_1))p + (1-\Phi(x_0))(1-p)}\\ &= \left(R(x_1)\frac{dx_0}{dy} - R(x_0)\frac{dx_1}{dy}\right)\\ &\quad \cdot \frac{p(1-p)\phi(x_1)\phi(x_0)}{((1-\Phi(x_1))p + (1-\Phi(x_0))(1-p))^2}.\end{aligned}$$

To simplify, calculate the derivatives

$$\begin{aligned}\frac{dx_1}{dy} &= \left(\frac{1}{2}+\sigma_y^2\ln\left(\frac{1-p}{p}\frac{t}{G(y)-t}\right)\right)\frac{d\sigma_y/dy}{\sigma_y^2}-\frac{\sigma_y G'(y)}{G(y)-t}\\ &= x_0\frac{d\sigma_y/dy}{\sigma_y}-\frac{\sigma_y G'(y)}{G(y)-t},\\ \frac{dx_0}{dy} &= \left(-\frac{1}{2}+\sigma_y^2\ln\left(\frac{1-p}{p}\frac{t}{G(y)-t}\right)\right)\frac{d\sigma_y/dy}{\sigma_y^2}-\frac{\sigma_y G'(y)}{G(y)-t}\\ &= x_1\frac{d\sigma_y/dy}{\sigma_y}-\frac{\sigma_y G'(y)}{G(y)-t}.\end{aligned}$$

Substituting and simplifying yields the form of the lemma. The inequality on K follows from $R'(x)<0$. □

Proposition 3. *Closer investments are more likely to succeed (i.e., $dT_y/dy<0$).*

Proof. From the product rule,

$$\frac{dT_y}{dy}=G(y)\frac{dH_y}{dy}-G'(y)H_y.$$

Substituting dH_y/dy from Lemma 4 and rearranging, $dT_y/dy<0$ if and only if

$$\begin{aligned}&(x_1R(x_1)-x_0R(x_0))\frac{d\sigma_y}{dy}\\ &<-G'(y)\sigma_y^2R(x_1)R(x_0)\left(\frac{1/\sigma_y}{G(y)-G(y)H_y}-\frac{1/R(x_0)-1/R(x_1)}{G(y)-t^*}\right).\end{aligned}$$

Since $xR(x)$ is increasing (from Lemma 2), the LHS is always negative. Note that $G(y)H_y\geq t^*$ by definition. Thus, the RHS is positive if

$$\frac{1}{R(x_0)}<\frac{1}{\sigma_y}+\frac{1}{R(x_0-1/\sigma_y).}$$

This follows immediately from Lemma 3. □

(Sub-)Model of Effort

The purpose of this section is to show that explicitly modeling effort yields an *equilibrium* probability that an undertaken investment succeeds, specified by $G(y)$, such that $G'(y)\leq 0$.

A successful venture investment yields profit π_V to a VC and π_C to a company. An unsuccessful venture yields profits normalized to 0 or to both. Let the probability that the venture succeeds be given by a function of VC and company effort, $\hat{G}(e_V, e_C)$, increasing, concave and complementary (i.e., $\hat{G}_{12}\geq 0$, or doing one's own job is more effective if the other party has done his or hers). The costs of effort for VC and company are given by a common, positive distance scalar, y, times respective, increasing, convex cost functions $C_V(e)$ and $C_C(e)$. Increasing cost of effort in distance may reasonably be thought of as higher communication costs between socially distant individuals or a host of other similar factors.

The VC and company solve

$$\max_{e_V}\{\pi_V\hat{G}(e_V, e_C^*)-yC_V(e_V)\}$$

and

$$\max_{e_C}\{\pi_C\hat{G}(e_V^*, e_C)-yC_C(e_C)\},$$

respectively. These yield the standard first-order conditions for both VC and company, in which marginal expected benefit equals marginal cost

$$\pi_V\hat{G}_1(e_V^*, e_C^*)=yC_V'(e_V^*)$$

and

$$\pi_C\hat{G}_2(e_V^*, e_C^*)=yC_C'(e_C^*),$$

respectively. Now we will show that effort for both parties decreases in distance y.

Proposition 4. (i) $de_V^*/dy<0$. (ii) $de_C^*/dy<0$.

Proof. Recall from the implicit function theorem that

$$\begin{aligned}\begin{pmatrix}\dfrac{de_V^*}{dy}\\[2ex] \dfrac{de_C^*}{dy}\end{pmatrix} &= -\begin{pmatrix}\dfrac{\partial f_V}{\partial e_V^*} & \dfrac{\partial f_V}{\partial e_C^*}\\[2ex] \dfrac{\partial f_C}{\partial e_V^*} & \dfrac{\partial f_C}{\partial e_C^*}\end{pmatrix}^{-1}\cdot\begin{pmatrix}\dfrac{\partial f_V}{\partial y}\\[2ex] \dfrac{\partial f_C}{\partial y}\end{pmatrix}\\ &= -\frac{1}{\Delta}\begin{pmatrix}\dfrac{\partial f_C}{\partial e_C^*}\dfrac{\partial f_V}{\partial y}-\dfrac{\partial f_V}{\partial e_C^*}\dfrac{\partial f_C}{\partial y}\\[2ex] -\dfrac{\partial f_C}{\partial e_V^*}\dfrac{\partial f_V}{\partial y}+\dfrac{\partial f_V}{\partial e_V^*}\dfrac{\partial f_C}{\partial y}\end{pmatrix},\end{aligned}\tag{11}$$

where the determinant of the Jacobian matrix is given by

$$\Delta=\frac{\partial f_C}{\partial e_C^*}\frac{\partial f_V}{\partial e_V^*}-\frac{\partial f_C}{\partial e_V^*}\frac{\partial f_V}{\partial e_C^*},\tag{12}$$

and the elements of the matrices are given by

$$\begin{aligned}\frac{\partial f_V}{\partial e_V^*} &= \frac{\partial}{\partial e_V^*}(\pi_V\hat{G}_1(e_V^*, e_C^*)-yC_V'(e_V^*))\\ &= \pi_V\hat{G}_{11}(e_V^*, e_C^*)-yC_V''(e_V^*)<0, &(13)\\ \frac{\partial f_V}{\partial e_C^*} &= \frac{\partial}{\partial e_C^*}(\pi_V\hat{G}_1(e_V^*, e_C^*)-yC_V'(e_V^*))\\ &= \pi_V\hat{G}_{12}(e_V^*, e_C^*)>0, &(14)\\ \frac{\partial f_C}{\partial e_V^*} &= \frac{\partial}{\partial e_V^*}(\pi_C\hat{G}_2(e_V^*, e_C^*)-yC_C'(e_C^*))\\ &= \pi_C\hat{G}_{12}(e_V^*, e_C^*)>0, &(15)\\ \frac{\partial f_C}{\partial e_C^*} &= \frac{\partial}{\partial e_C^*}(\pi_C\hat{G}_2(e_V^*, e_C^*)-yC_C'(e_C^*))\\ &= \pi_C\hat{G}_{22}(e_V^*, e_C^*)-yC_C''(e_C^*)<0, &(16)\\ \frac{\partial f_V}{\partial y} &= \frac{\partial}{\partial y}(\pi_V\hat{G}_1(e_V^*, e_C^*)-yC_V'(e_V^*))\\ &= -C_V'(e_V^*)<0, &(17)\\ \frac{\partial f_C}{\partial y} &= \frac{\partial}{\partial y}(\pi_C\hat{G}_2(e_V^*, e_C^*)-yC_C'(e_C^*))\\ &= -C_C'(e_C^*)<0. &(18)\end{aligned}$$

All inequalities follow because G is increasing, concave, and complementary, and C_i is increasing and convex.

We claim that the determinant of the Jacobian is positive (i.e., $\Delta>0$). Expanding the terms in (12) according to their definitions in Equations (13)–(18), this is true if and only if

$$\begin{aligned}&(\pi_C\hat{G}_{22}(e_V^*, e_C^*)-yC_C''(e_C^*))(\pi_V\hat{G}_{11}(e_V^*, e_C^*)-yC_V''(e_V^*))\\ &\quad>\pi_V\pi_C\hat{G}_{12}(e_V^*, e_C^*)^2,\end{aligned}$$

which can be rearranged:

$$\pi_V \pi_C(\hat{G}_{11}(e_V^*, e_C^*)\hat{G}_{22}(e_V^*, e_C^*) - \hat{G}_{12}(e_V^*, e_C^*)^2)$$
$$> \pi_V \hat{G}_{11}(e_V^*, e_C^*) y C_C''(e_C^*) + \pi_C \hat{G}_{22}(e_V^*, e_C^*) y C_V''(e_V^*)$$
$$- y^2 C_V''(e_V^*) C_C''(e_C^*).$$

The LHS is positive because of the concavity of G (i.e., $\hat{G}_{11}\hat{G}_{22} - \hat{G}_{12}^2 > 0$), and the RHS is negative because of the concavity of G (i.e., $\hat{G}_{11} < 0$ and $\hat{G}_{22} < 0$) and the convexity of C (i.e., $C_i'' > 0$). Thus, we have proved our claim that $\Delta > 0$.

Finally, then, the proposition holds if and only if the elements of the computed column vector in Equation (11) are positive. Both elements sign positively directly from applying the inequalities in Equations (13)–(18). □

Thus, smaller distances increase effort by both parties. Since G increases in both parameters, we know that the probability of a successful venture decreases as distance grows. In other words, writing the equilibrium efforts as functions of distance $\hat{G}(e_V^*, e_C^*) = G(y)$, where $G'(y) \le 0$, as specified in the main model, is a sufficient statistic for an embedded effort model.

One might model other unobserved features in which postinvestment success, such as monitoring effectiveness or coordination, depend on (ethnic) distance to similar effect. The details of the various extensions would differ slightly, but the comparative static on the equilibrium probability of success with respect to distance would also be negative—our current formulation captures this essential feature generally. The basic intuition is that if social distance interferes with one of the activities positively related to the success of the venture or makes it more costly to engage in, the less the activity will occur and the probability of success will decrease.

References

Ackerberg D, Botticini M (2002) Endogenous matching and the empirical determinants of contract form. *J. Political Econom.* 110(3):564–591.

Agrawal A, Kapur D, McHale J (2008) How do spatial and social proximity influence knowledge flows? Evidence from patent data. *J. Urban Econom.* 64(2):258–269.

Bengtsson O, Hsu DH (2013) Ethnic matching in the U.S. venture capital market. Working paper, The Wharton School, University of Pennsylvania, Phildelphia.

Bergemann D, Hege U (1998) Venture capital financing, moral hazard, and learning. *J. Banking Finance* 22(6–8):703–735.

Bottazzi L, Da Rin M, Hellmann T (2008) Who are the active investors? Evidence from venture capital. *J. Financial Econom.* 89(3):488–512.

Bottazzi L, Da Rin M, Hellmann T (2012) The importance of trust for investment: Evidence from venture capital. Working paper, University of British Columbia, Vancouver.

Bound J, Jaeger DA, Regina B (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *J. Amer. Statist. Association* 90(430):443–450.

Bratter JL, King RB (2008) But will it last? Marital instability among interracial and same-race couples. *Family Relations* 57(2):160–171.

Cochrane J (2006) The risk and return of venture capital. *J. Financial Econom.* 75(1):3–52.

Fisman R, Paravasini D, Vig V (2012) Cultural proximity and loan outcomes. NBER Working Paper 18096, National Bureau of Economic Research, Cambridge, MA.

Gompers PA, Lerner J (2004) *The Venture Capital Cycle* (MIT Press, Cambridge, MA).

Gompers PA, Mukharlyamov V, Xuan Y (2012) The cost of friendship. NBER Working Paper 18141, National Bureau of Economic Research, Cambridge, MA.

Gordon RD (1941) Values of Mill's ratio of area to bounding ordinate of the normal probability integral for large values of the argument. *Ann. Math. Statist.* 12:364–366.

Gould DM (1994) Immigrant links to the home country: Empirical implications for U.S. bilateral trade flows. *Rev. Econom. Statist.* 76(2):302–316.

Greene W (2001) Estimating econometric models with fixed effects. Working paper, Stern School of Business, New York University, New York.

Heckman J (1979) Sample selection bias as a specification error. *Econometrica* 47(1):591–530.

Heckman J (1981) The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. Manski C, McFadden D, eds. *Structural Analysis of Discrete Data with Econometric Applications* (MIT Press, Cambridge, MA), 179–196.

Hellmann T (2000) Venture capitalists: The coaches of Silicon Valley. Miller W, Lee CM, Hanock MG, Rowen H, eds. *The Silicon Valley Edge: A Habitat for Innovation and Entrepreneurship* (Stanford University Press, Stanford, CA), 267–294.

Hochberg Y, Ljungqvist A, Lu Y (2007) Whom you know matters: Venture capital networks and investment performance. *J. Finance* 62(1):266–301.

Kalnins A, Chung W (2006) Social capital, geography, and survival: Gujarati immigrant entrepreneurs in the U.S. lodging industry. *Management Sci.* 52(2):233–247.

Kaplan SN, Strömberg P, Sensoy BA (2002) How well do venture capital databases reflect actual investments? Working paper, University of Chicago, Chicago.

Kerr W (2008) Ethnic scientific communities and international technology diffusion. *Rev. Econom. Statist.* 90(3):518–537.

Lerner J (1995) Venture capitalists and the oversight of private firms. *J. Finance* 50(1):301–318.

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

Morgan J, Várdy F (2009) Diversity in the workplace. *Amer. Econom. Rev.* 99(1):472–485.

National Venture Capital Association (2009) *Venture Impact: The Economic Importance of Venture Capital-Backed Companies to the U.S. Economy*, 5th ed. (National Venture Capital Association, Arlington, VA).

Rider CI (2012) How employees' prior affiliations constrain organizational network change: A study of U.S. venture capital and private equity. *Admin. Sci. Quart.* 57(3):453–483.

Rosenbaum PR, Rubin DB (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Amer. Statist.* 39(1):33–38.

Sampford MR (1953) Some inequalities on Mill's ratio and related functions. *Ann. Math. Statist.* 24(1):130–132.

Sinnott RW (1984) The virtues of the Haversine. *Sky Telescope* 68(2):158.

Sørensen M (2007) How smart is smart money? An empirical two-sided matching model of venture capital. *J. Finance* 62(6): 2725–2762.

Sorenson O, Stuart T (2001) Syndication networks and the spatial distribution of venture capital investments. *Amer. J. Sociol.* 106(6):1546–1588.

Stock J, Yogo M (2005) Testing for weak instruments in IV regression. Andrews DWK, Stock JH, eds. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (Cambridge University Press, Cambridge, UK), 80–108.

Tian X (2011) The causes and consequences of venture capital stage financing. *J. Financial Econom.* 101(1):132–159.

Wang S, Zhou H (2004) Staged financing in venture capital: Moral hazard and risks. *J. Corporate Finance* 10(1):131–155.

Webber R (2007) Using names to segment customers by cultural, ethnic or religious origin. *J. Direct Data Digital Marketing Practice* 8(3):226–242.