Due Diligence and the Allocation of Venture Capital

Jack Xiaoyong Fu
Lucian A. Taylor*

August 29, 2025

Abstract

How do investors choose the intensity of their due diligence, and how does that choice affect investment outcomes? Using smartphone signal data, we measure the duration of pre-investment meetings between venture capitalists (VCs) and startup employees. This measure captures one important component of VC due diligence. Less due diligence is associated with hotter deals and markets, busier investors, and greater distance, consistent with a theory of costly learning. Also consistent with that theory, less due diligence is associated with more volatile investment performance, as VCs allocate capital under greater uncertainty. Overall, VCs appear to trade off the costs of due diligence with its improvements to capital allocation.

JEL classifications: E22, G11, G24
Keywords: Due diligence, venture capital, capital allocation

*Fu is at the University of Hong Kong (jackfxy@hku.hk). Taylor is at the Wharton School of the University of Pennsylvania and the NBER (luket@wharton.upenn.edu). We are grateful for comments from Mike Ewens, Xavier Giroud, and Tom Winberry; participants at the 2025 NBER Summer Institute (Big Data and High-Performance Computing); and seminar participants at the following universities: American, Arizona State, Columbia, and Pennsylvania. We thank Yiwen Lu and Chris Yaoyao Zhu for excellent research assistance. This research was supported by the Mack Institute and Jacobs Levy Equity Management Center. The Institutional Review Board at the University of Pennsylvania determined that the research meets the eligibility criteria for IRB review exemption authorized by 45 CFR 46.104, category 4.

1. Introduction

Due diligence is widespread in practice but largely absent from empirical research. Due diligence refers to the process of evaluating an asset prior to investing in it. In VC, due diligence is the primary way investors select which startups to fund, how much to invest, and on what terms. Selection, in turn, is ranked by VCs as the most important source of value creation—even more important than deal flow and post-investment support (Gompers et al., 2020). Because VC-backed firms play a large role in innovation, diligence decisions affect not only fund returns but also the allocation of capital to new ideas, technologies, and markets. Diligence entails significant costs, with VCs spending an average of 118 hours on due diligence per investment (Gompers et al., 2020). VCs cite these efforts when justifying the high fees they charge their limited partners (e.g., Wilson, 2006).

We study two related questions: How do VCs choose how much due diligence to perform, and how does that choice affect investment outcomes? While practitioners emphasize that diligence is costly and important, existing evidence is largely indirect and often mixed. Without direct measures of diligence intensity, it has been difficult to document when and why diligence varies, or to assess whether it improves capital allocation.

We address this gap using a new, large-scale, objective proxy for VC due diligence: the total time VC and startup employees spend together in person in the months leading up to a financing round. We construct this measure from smartphone geolocation data and link it to deal-level information from PitchBook, covering roughly 22,000 U.S. VC deals from 2018 to 2023. This approach allows us to capture pre-investment, in-person interactions at scale without relying on self-reports or single-investor case studies.

Our analysis yields two main messages. First, diligence intensity is a choice variable that responds to a variety of costs. In our data, VCs perform less diligence when deals, sectors, or markets are "hot," when VCs are busier with other investments, and when startups are geographically distant. The diligence choice therefore depends not only on startup characteristics but also on investor characteristics and market conditions. Second, due diligence helps VCs allocate capital efficiently. More diligence is associated with less dispersion in the marginal product of capital (MPK) across deals, consistent with a model in which pre-investment learning reduces uncertainty and improves investment choices. Value creation is quite sensitive to diligence levels. Together, these results imply that pre-investment effort is an important channel through which VCs create value.

We focus on one important component of VC due diligence: in-person meetings between investors and startup employees. We identify these meetings using anonymized, highly granular smartphone geolocation data. Devices likely belonging to VC or startup employees are identified based on recurring presence at known office addresses, and meeting duration is measured when VC employees visit startup offices or vice versa. Our proxy for due diligence intensity is the total number of hours spent in these meetings during the 18 months before an investment. While most of our investors are traditional VCs, the sample also includes accelerators, incubators, corporate venture capital (CVC) investors, asset managers, and other nontraditional investors. For convenience, we refer to all these investors as "VCs."

We test how due diligence responds to a variety of costs. Several of our proxies relate to how "hot" the given deal, sector, or overall VC market is. We argue that due diligence costs more in hotter markets or deals, for two potential reasons. On the demand side, if a deal or market is hot because there are many attractive opportunities, then investors are busier, and the opportunity cost of their time is higher. On the supply side, if a market is crowded with many investors, then extra diligence imposes an indirect cost by increasing the chance that a competing investor steals the deal. To measure how hot an individual deal is, we count the number of other VCs—besides the one who ultimately makes the investment—that meet with the startup in the period leading up to the investment. We measure how hot a sector is by the number of VC deals that occur in the same year, stage, and industry relative to previous years. Finally, to measure how hot the overall VC market is, we simply count the VC deals per quarter.

We find less due diligence when the deal, sector, or VC market is hotter. VC deal volume and due diligence have aggregate time-series correlations ranging from -33% to -85% across specifications. Despite our short time series, these correlations are statistically significant in most specifications, and they are especially strong in the post-COVID period, after 2020. In deal-level regressions, the amount of diligence is negatively related to how hot both the sector and individual deal are. A one standard deviation increase in abnormal deal volume at the sector level is associated with a 15%–34% reduction in diligence hours. Doubling the number of other VCs meeting with the startup is associated with a 13% decrease in diligence hours. These results are robust to including time fixed effects, which sweep out effects such as COVID-19 and the rise of virtual meetings, as well as stage fixed effects, which account for potentially different levels and styles of due diligence across stages.

To more directly measure how busy investors are, we compute the ratio of the number of recent deals to the number of investment professionals at the VC firm. This ratio has a strongly negative relation to our diligence measure. For example, even in the presence of industry-by-month fixed effects, the negative relation is significant at the 1% confidence level and implies a 22% decrease in diligence levels in relation to a doubling of deals per employee. In short, busier investors perform less due diligence.

Another cost of due diligence relates to travel time. Our diligence measure is significantly lower when the investor and startup are farther apart geographically, consistent with higher travel costs reducing in-person due diligence.

To help interpret these facts and generate testable predictions about investment outcomes, we provide a simple model. The model features an investor who chooses how much to learn about a startup before investing. Learning is costly but helps the investor choose how much to invest. Consistent with the facts above, the model predicts that when learning is more costly, investors optimally choose to learn less, which we interpret as performing less due diligence.

According to the model, less due diligence is associated with a higher variance of deal performance, as measured by the investment's MPK. MPK in our model corresponds to the startup's valuation step-up per dollar invested. Intuitively, the less a VC learns through due diligence, the more likely the VC is to either over- or under-invest in the startup, producing either very low or very high investment performance—i.e., higher dispersion.

Consistent with this prediction, we show empirically that lower levels of due diligence are associated with higher dispersion in investment outcomes. Guided by the model, we measure a VC investment's MPK as the ratio of (1) the startup's valuation step-up from this financing round to the next, to (2) the amount invested in this financing round. We model the variance of MPK using the squared deviation of MPK from its expected value, deal by deal. In deal-level regressions, we find a significantly negative relation between the variance of MPK and our diligence measure. This result holds even in the presence of month, industry, and stage fixed effects. We also find a negative relation between our diligence measure and the variance of VC returns, which are highly correlated with investments' MPKs. In the aggregate time series, there is more MPK dispersion during hot VC markets, consistent with reduced due diligence during those periods. We also find higher MPK dispersion among younger startups and investors with less industry-specific experience, consistent with dispersion capturing investor uncertainty.

According to the model, the variance of MPK also captures the degree of capital misallocation. This idea originates from Hsieh and Klenow (2009) and the large macroeconomics literature that follows it. Many papers in that literature quantify resource misallocation using dispersion in the marginal products of production inputs. We follow suit, recognizing that "misallocation" in our setting is ex post and arises not from investor mistakes but from imperfect information. Interpreted through our model, the estimated negative relation between diligence intensity and MPK dispersion implies that additional due diligence improves the allocation of venture capital to startups.

Our model allows us to quantify this effect. By feeding our estimated regression coefficients into the model, we measure how a hypothetical doubling of hours spent on due diligence would change the expected amount of value created in a VC deal. In our baseline calibration, we find this improvement to be 6% of the VC's amount invested. The improvement approaches 20% in alternative calibrations. These calibrations are simple and omit many features of reality, but they do suggest that due diligence plays a large role in value creation and the efficient allocation of capital to startups.

The regressions above relate due diligence to dispersion in investment performance. Does due diligence also relate to the level of performance? The model predicts no relation between the amount of diligence and the deal's expected MPK. In the model, less diligence leads investors to over- or under-invest more often. However, investors are rational learners and, therefore, get it right on average, so diligence intensity does not affect a deal's expected MPK. Consistent with this prediction, in most regression specifications we find no significant relation between our diligence measure and either the level of MPK or the level of return.

We extend the analysis to include due diligence on rejected deals. We find that diligence patterns across all deals—accepted and rejected—help predict startup quality, proxied by whether the startup reaches an exit or raises a subsequent round. Among deals not funded by a top VC, startup quality increases with the amount of due diligence performed by top VCs. Intuitively, conditional on being rejected, receiving more attention from top VCs is a positive signal. The reverse is true for startups that do get funded by a top VC: startup quality decreases in the amount of diligence, arguably because very positive signals allow the VC to terminate diligence early. Obtaining funding from top VCs predicts higher startup quality, consistent with the findings of Sørensen (2007) and Ewens and Rhodes-Kropf (2015). We extend that literature by showing that diligence patterns themselves provide extra predictive power. Our findings suggest that the matching between VCs and startups on quality begins as early as the diligence stage, and that top VCs deploy their diligence efforts strategically.

Since our diligence measure is new, we perform several validity checks. In deals where we can identify at least one meeting, the average amount of due diligence we measure is 32 hours, and the median is much lower. For comparison, the survey of Gompers et al. (2020) finds that VCs spend an average of 118 hours on diligence per deal, so clearly we measure only a fraction of all diligence activities. Consistent with their survey, we find less due diligence by

VCs in California and VCs in the tech sector. Unlike Gompers et al., we do not find more diligence in later-stage deals, possibly because the in-person meetings we measure make up a smaller portion of total diligence activities at later stages. If we use our smartphone data to estimate the number of startups a VC meets with per finalized investment or, conversely, the number of VCs a startup meets with while raising a financing round, we find magnitudes quite similar to those in existing surveys.

These validity checks give some comfort, but our diligence measure clearly has limitations. It misses virtual meetings and phone calls, and it misses diligence activities that do not require a meeting. It misses meetings if certain apps are not running in the background on a user's phone. These measurement issues clearly introduce noise, but this noise does not necessarily bias our tests. We argue that in almost all cases, measurement error either introduces no bias or introduces bias that works against our conclusions. There is clearly more work to do on measuring investor due diligence, but we hope our study takes a useful step forward on this important, under-researched topic.

There is indeed little empirical research on VC due diligence. Early descriptions of the topics and activities involved in VC due diligence come from Tyebjee and Bruno (1984) and Kaplan and Stromberg (2001, 2004). More recently, Gompers et al. (2020) survey almost 900 VCs and provide statistics on their process for making investment decisions, including the amount of time and the actions involved in due diligence. Jang and Kaplan (2025) analyze data on deals sourced and scored by one VC, finding that the VC has selection ability and places more weight on team characteristics when making initial investment decisions. Indirectly related to due diligence, several papers relate startup and deal characteristics to either investor interest or investment outcomes.[1] Even outside VC, there is little empirical research on due diligence.[2]

The COVID-19 episode offers mixed evidence on the importance of in-person VC due diligence. COVID-19 limited in-person meetings, yet VCs surveyed during the pandemic say they did not significantly change their time allocations or struggle to find good entrepreneurs (Gompers et al., 2021). Those results suggest only a minor role for in-person diligence. Three facts point to a larger role, however. VCs reported difficulty evaluating deals during COVID (Gompers et al., 2021), perhaps due to limited in-person meetings. Alekseeva et al. (2025)

[1] See, e.g., Baum and Silverman (2004), Gompers et al. (2010), and Bernstein et al. (2017). Like us, Lyonnet and Stern (2022) study the allocative efficiency of VC, albeit with a focus on machine learning.

[2] Gompers et al. (2016) survey 79 private equity (PE) investors and describe their deal-selection process and criteria. Cumming and Zambelli (2017) study PE due diligence using a survey of Italian investors. Brown et al. (2023) analyze a single institution's diligence of hedge fund managers. Offenberg and Pirinsky (2015) and Wangerin (2019) study due diligence in M&A deals.

study VCs' adaptation to COVID-19 and conclude that online interactions are not a perfect substitute for in-person meetings. Surveyed VCs report higher dispersion in returns during COVID-19 (Gompers et al., 2021), consistent with the link we find between low due diligence and more-dispersed investment outcomes.

More broadly, the idea that investors engage in costly learning dates back at least to Grossman and Stiglitz (1980). Since then, a large theory literature has studied investor learning. In VC, several facts point to the importance of this learning and a desire to reduce its costs: VCs tend to locate in tech hubs and invest locally (Chen et al., 2010; Bernstein et al., 2016); VCs and other investors often invest in socially connected firms (Kuchler et al., 2022; Hochberg et al., 2007); and VCs have begun hiring data scientists to help screen startups (Bonelli, 2025).

The existing evidence is useful but has limitations. Surveys and case studies provide granular evidence but raise concerns about external validity. The observational studies do not measure information-gathering activities, so their evidence is indirect. We contribute to this literature by directly measuring VCs' information-gathering activities in a large sample of investors and deals. Doing so allows us to empirically relate the amount of diligence to market conditions, investment outcomes, and startup quality. Directly relating the amount of diligence to other costs, such as investor busyness and distance, also seems new.

This paper also belongs to the large literature on capital misallocation, following Hsieh and Klenow (2009).[3] Within this literature, other papers that relate misallocation to uncertainty and imperfect information include Asker et al. (2014), David et al. (2016), David and Venkateswaran (2019), David et al. (2022), and Charoenwong et al. (2024). To our knowledge, estimating misallocation in VC or relating it to due diligence is new.

Our due diligence proxy relates to Fu (2024), who uses the same smartphone signal data to measure meetings between VCs and startups. However, Fu (2024) studies a different topic: post-investment VC monitoring and its reputational effects. Using smartphone geolocation data in economics is relatively new.[4]

The remainder of the paper is organized as follows. Section 2 describes our data and institutional details on VC due diligence. Section 3 provides stylized facts and validity checks regarding our measure. Section 4 relates the chosen amount of due diligence to its costs. Section 5 presents our model of the diligence choice and its implications for capital

[3]Recent applications in finance include Cong et al. (2019), Ai et al. (2020), Whited and Zhao (2021), Catherine et al. (2022), and Choi et al. (2025).

[4]Other papers using smartphone geolocation data include Chen et al. (2022), Chen et al. (2023), Atkin et al. (2025), and Gerken et al. (2025).

allocation and investment outcomes. Section 6 empirically tests those implications and provides a simple quantification. Section 7 extends the analysis to include rejected deals. Section 8 explores robustness, and Section 9 concludes.

2. Data and institutional background

2.1. Due diligence in VC

The goal of VC due diligence is to assess a startup's potential and verify its claims. During due diligence, investors evaluate the management team, product, technology, market size, competitive landscape, business model, valuation, legal status, and other topics.

VC due diligence takes a variety of forms and can last anywhere from a few days to over a year (Gompers et al., 2020). The process begins after a member of the VC firm sources a potential investment opportunity. The VCs then typically hold informal and then formal pitch meetings with the startup's management team. We interpret these meetings as the first steps in the due diligence process, as the investor is beginning to evaluate the startup. After those meetings, a period of formal due diligence begins. Activities in that period include more meetings between VCs and startup employees, reference checking, consulting customers and external experts, data gathering and analysis, and financial modeling. If the VCs are satisfied with what they learn during diligence, they offer the startup a term sheet, which proposes an investment amount in exchange for cash-flow and control rights. The term sheet is mostly non-binding and is followed by more due diligence. If that goes well, then the final, potentially revised deal terms are drafted, and the deal closes.

On average, roughly 100 startups begin the due diligence process per one investment made, and the process can end at various points along the way (Gompers et al., 2020). The typical process involves extensive contact between the investors and startup employees. In some cases, though, an investor will conduct an abbreviated due diligence without the startup's knowledge or any meetings, and the investor will then offer an unsolicited term sheet, a strategy known as "round pre-emption" (Plapperer, 2022).

2.2. Measuring due diligence

We compute our due diligence proxy using data on smartphone signals near VC and startup office buildings. We obtain these data from a leading smartphone data vendor. Smartphone operating systems (Android and iOS) record the longitude and latitude of a phone with timestamps every 5 to 10 minutes, and more frequently when the user is driving. These location estimates can be accurate within 20 meters and, subject to user permissions, are shared with apps that are open or running in the background. The data vendor collects this location data from hundreds of popular apps covering multiple categories. The vendor's data coverage is quite good, with 220 to 240 million monthly active users in the U.S., which represent roughly 80% of all smartphones.[5]

Using PitchBook data, we list the addresses of all VC and startup office buildings in the dataset. The vendor then provides us the data on all smartphone signals near those addresses. The dataset is constructed in two main phases. First, we identify devices likely belonging to VC employees. To differentiate VC employees from passersby, we examine the frequency of a device's presence near the VC office. A device detected within 200 meters of the VC office on at least five working days per month, across two months, is flagged as likely belonging to an employee. To further differentiate VC employees from frequent visitors like delivery workers, we exclude devices flagged as employees at more than five companies. We apply the same procedure to identify startup employees.

Second, we detect potential meetings between VCs and startups. If a potential VC employee device is detected within 200 meters of a startup's office and remains there for at least 10 minutes, we consider it a potential meeting. If multiple VC employees visit the startup building on the same day, we take the maximum duration as the meeting time.

We apply several additional filters to mitigate false positives (instances where no actual meeting occurs, but we mistakenly consider it one). First, meetings lasting longer than five hours are flagged as false positives since they likely indicate other activities. Second, if a VC visits a startup more than 10 times in a single month, it is also considered a false positive, possibly due to mistaking passersby for employees. Third, we only count meeting hours when at least three smartphone signals are captured in the interval, as a higher number of signals indicates a higher likelihood that the VC is continuously staying around the building. Lastly, we focus on interactions occurring within 18 months prior to the investment date, ignoring those beyond this time frame. We use the same methods to measure startup employees visiting VC buildings.

We use these data to compute our proxy for due diligence intensity: the total number of hours that investors and startup employees spend together, either at the startup or investor's

[5] According to a 2023 Pew Research Center report (see `https://www.pewresearch.org/internet/fact-sheet/mobile/`), 90% of U.S. adults report owning a smartphone, and the U.S. population is approximately 335 million in 2023, so the number of smartphone users is approximately 300 million.

building, within 18 months prior to the investment date.

By focusing on in-person meetings, our proxy captures a critical part of VC due diligence. The proxy has limitations, however. We miss phone calls and virtual meetings, an issue we discuss below. We also miss in-person meetings that occur in alternative locations, such as restaurants. We miss in-person meetings when smartphones are turned off, lose reception, or have relevant apps neither open nor running in the background. This source of measurement error, which is determined by individual smartphone usage habits (such as how often users clear background apps), is unlikely to correlate with the variables we study. Our proxy also omits diligence activities that do not involve meetings between VC and startup employees, such as reference-checking, financial modeling, and consulting customers and external experts.

Clearly, our results pertain to only a subset of due diligence activities. If investors scale these various diligence activities up or down proportionally, then our proxy correlates perfectly with total due diligence. To the extent that investors do not scale these activities proportionally, we measure total diligence intensity with error. We discuss potential bias from unmeasured or mismeasured due diligence below, when interpreting the regression results. Section 8 further explores the robustness of our diligence measure.

2.3. Sample formation

To reduce measurement error in our due diligence proxy, we focus on diligence in a sample of deals that reach completion. If we know VC A actually invested in startup B, then the meetings we measure leading up to that known deal are more likely to reflect true meetings between employees of A and B rather than false positives.

To build our sample, we start with all PitchBook deals from January 2018 to January 2023 that have valid information for the deal date and lead investor, resulting in approximately 403,000 investor-startup observations in 174,000 deals. We apply several filters. We require that investors are headquartered in the U.S. and have a valid headquarters address. We exclude investors categorized by PitchBook as "Corporation" (a distinct category from CVC in PitchBook), "PE-Backed Company," "VC-Backed Company," or "Other." The investors must also have active status and be listed in the PitchBook VC North America Data. Roughly 116,000 observations remain after applying these filters. Next, we require startups to have headquarters in the U.S. with a valid address (roughly 104,000 observations remaining). Last, we apply deal-specific filters. We require that the investor and startup have different office addresses (about 103,000 observations remaining). The deal type must be "Early Stage VC," "Later Stage VC," "Seed Round," or "Accelerator/Incubator" (about 91,000 observations remaining). Because nontraditional investors (e.g., buyout firms, asset managers) also participate in these rounds, our sample includes them as well. We then require the investor to be the lead investor (about 26,000 observations remaining). We focus on lead investors because non-lead investors typically conduct little to no diligence. We also require the investor to be a new investor in the startup, because otherwise we could not distinguish between monitoring a previous round's investment and doing diligence on the next round. Our final sample covers approximately 21,000 deals.

3. Stylized facts and validation

Table 1 contains summary statistics on the due diligence measure and other variables that we introduce later. Appendix A.1 contains detailed definitions of all variables. In the full sample, the average due diligence measure is 1.5 hours, but the measure equals zero in 95% of observations. A zero value indicates either there was no in-person meeting, or there was a meeting but the smartphone data fail to capture it, e.g., because phones were turned off or did not have relevant apps running in the background. Later, in any analysis that includes the zeros, we show an alternative version of the analysis that excludes the zeros.

A case can be made for either including or excluding the zeroes. On the one hand, the zeroes may reflect particularly severe measurement error, such as a wrong address, and therefore should be excluded. On the other hand, some VCs truly invest without an in-person meeting (see Section 2.1), in which case the zeroes are accurate. Even if measurement error leads to some zero values, a zero value is still informative in the sense that it indicates that the true, total, unobservable number of meetings is likely to be very low for that deal.

Panel B of Table 1 shows summary statistics for the subsample of deals with a positive due diligence value. There is high dispersion in the diligence measure, which ranges from 1.3 hours at the 25th percentile to 19.4 hours at the 75th percentile. Values in the far-right tail push the average much higher, to 32 hours. Later, we take logs to tame these outliers.

As a validity check, we compare our summary statistics to those in Gompers et al. (2020), which contains results from a survey of 885 VCs conducted in 2015–2016. They find that VCs on average spend 118 hours on due diligence per deal. Since we measure only a subset of diligence activities, it is comforting that the average we find (e.g., 32 hours in Panel B) is below their average. Comparing those averages suggests our measure captures approximately

27% (= 32/118) of total hours spent on due diligence. Of course, that percentage would be lower if we included the zero values, as we do in Panel A.

Table 2 compares due diligence levels across different subsamples. Panel A shows similar levels of due diligence across early-stage, later-stage, accelerator/incubator, and seed rounds. These differences push against the notion that investors often "spray and pray" in the earliest rounds, investing without much due diligence. The one statistically significant difference we find actually points in the opposite direction: a positive value of due diligence is actually more likely in seed rounds than in early-stage deals. Unlike us, Gompers et al. (2020) find significantly more total hours spent on due diligence in late-stage than early-stage deals. These findings can be reconciled if, as companies mature, a growing share of diligence shifts away from in-person meetings toward other activities such as financial modeling, customer calls, and data analysis. At the earliest stages, those avenues are limited—products and customers may not exist, and financials are too thin—so diligence naturally centers on the founding team (Gompers et al., 2020), best evaluated through in-person meetings. Given these stage-related differences, we include stage fixed effects in our main analyses.

Panel B compares startups' industries. Consistent with Gompers et al. (2020), we find more diligence in healthcare compared to IT deals. The differences are not statistically significant, but they are large in magnitude (e.g., 42 vs. 31 hours on average, conditioning on positive values). Diligence levels are strikingly similar between IT, B2C, B2B, financial services, and energy deals.

Panel C shows large geographic differences in diligence intensity. Compared to investors in California, those in New York, Boston, and "Other" are three to four times more likely to have a meeting recorded in our sample (i.e., a positive due diligence value). Conditional on having a recorded meeting, the average time VCs spend in meetings is 1.6–2.3 times greater outside California. Gompers et al. (2020) find a similar pattern, with 81 hours spent on diligence in California and 129 hours in other U.S. locations, on average. These results suggest interesting differences in investing styles between East Coast and West Coast VCs.

Panel D compares investor types. Investors classified as VCs make up by far the largest group. Compared to VCs, accelerator/incubator investors show similar levels of due diligence, which is surprising given their small deal sizes. Growth investors, CVCs, PE/buyout firms, and asset managers are considered "nontraditional" investors, a category that grew in prominence during the past decade (e.g., Ewens and Farre-Mensa, 2020; Chernenko et al., 2021). Compared to VCs, PE/buyout firms perform slightly more diligence, while CVCs perform slightly less. Those differences are not significantly different. Diligence levels, however, are significantly lower among growth/expansion investors and asset managers. Examples of the latter in our sample include Goldman Sachs Asset Management, T. Rowe Price, and Tiger Global. On the one hand, those investors' lower levels of diligence are surprising given their larger deal sizes. On the other hand, our results confirm anecdotes about certain nontraditional investors making deals faster and with less diligence during this time period.[6]

Figure 3 relates due diligence intensity to two investor characteristics related to scale. Each panel shows a binscatter plot of the log due diligence measure (on the y-axis) versus an investor characteristic (on the x-axis). Larger VC firms do less due diligence on average (Panel A), perhaps because their stronger reputations attract higher-quality deals, which require less scrutiny. In Panel B, we see a strong negative relation between diligence levels and the number of investments the VC makes within 18 months of the focal deal. One potential explanation is that VCs follow different investing strategies. Some VCs "spray and pray" (Ewens et al., 2018), making many investments with little due diligence. Other VCs follow a more selective strategy, making fewer investments but performing more diligence on each. Another potential explanation is that VCs do more deals and perform less diligence when markets are hotter. In either explanation, investors with more recent deals are busier, making due diligence more costly. We explore this pattern more in the next section.

For further validation, we use our smartphone data to compute each investor's ratio of (1) the number of startups with which it meets to (2) the number of investments it makes during the sample period.[7] We plot the distribution of this pitch-to-investment ratio across investors in Figure 1. The average (median) investor meets with 60 (14) startups per deal closed. For comparison, according to the survey evidence of Gompers et al. (2020), the average VC investor meets with 28 startups' management teams per deal closed. The similarity between these numbers suggests our method of using smartphone data to identify investor-startup meetings does not suffer from a major imbalance of false positives and false negatives. The distribution in Figure 1 is also interesting in its own right. Some investors appear very selective, meeting with over 100 startups per deal. Others are much less so, meeting with fewer than 10 startups per deal.

As a final, related validation, we analyze the number of VC firms each startup pitches to per funding round. We treat any pre-investment meeting between the startup and a VC investor as a pitch meeting. Figure 2 shows the distribution of pitches per funding round,

[6]For example, hedge fund Tiger Global gained a reputation in 2021 for "writing checks within mere days" and backing "the equivalent of nearly one startup every day—including weekends" (Mathews and Sraders, 2023).

[7]Whereas our main measure includes only diligence in completed deals, the measure used here also includes meetings with startups that do not receive the investor's money.

treating each round as one observation. These results can be compared to those from First Round Capital's "State of Startups," a 2017 survey of 869 venture-backed startup founders.[8] Similar to the numbers plotted in Figure 2, the survey asks founders how many investment firms they pitched to when raising their last round. We find that in 56% of rounds the startup pitched to fewer than 10 investors, whereas the survey finds 57%. We find that in 12% of rounds, the startup pitched to 11–20 investors, compared to 20% in the survey. The startup pitched to more than 20 investors in 32% of our rounds, compared to 23% in the survey. These numbers are all quite close, again suggesting that our method for using smartphone data to identify startup-investor meetings is reasonable.

4. The due diligence choice

Whether to perform more or less due diligence is an investor choice influenced by a variety of costs and constraints imposed by the startup and external factors. We formally model that choice in Section 5. One simple prediction emerging from our model is that VCs do less due diligence when it is more costly. We test that prediction in this section. We consider costs related to travel time, being busy, and whether markets are hot or cold.

Our simplest tests involve aggregate time-series correlations. We predict less due diligence when the VC market is "hotter," i.e., when more deals are being made. Due diligence becomes more costly in hot markets for three reasons. First, in hot markets there is a higher risk of having a deal picked off by another investor during the diligence period, pushing investors to move fast.[9] Second, the opportunity cost of a VC's time is likely higher if hot markets feature more and better investment opportunities. Third, if bargaining power shifts toward startups during hot markets, and if startups prefer shorter diligence, then lengthy diligence in a hot market makes the startup more likely to reject the VC. Pushing against these predictions, if hot markets feature startups with more soft information, and if soft information is best assessed in-person, then we would expect more meetings in hot markets.

At the VC industry level, we find a negative relation between due diligence levels and VC deal volume, as seen in Figure 4. That figure plots the number of VC deals in our sample each quarter and the median diligence length across those deals. The two series' correlation is −33%. Diligence levels are especially low from 2021q2 through 2022q1, after deal volume

[8] See `https://stateofstartups.firstround.com/2017/`.

[9] VC term sheets typically include an exclusivity period, which allows the VC to conduct final due diligence without the risk of being picked off. Most diligence, however, occurs before the term sheet is signed, when that risk still remains.

exploded. Diligence levels then increase sharply in 2022q3, after deal volume crashed. These patterns agree with anecdotes about investors cutting back on diligence during the hot period of 2021, then extending their diligence once the market cooled in 2022.[10]

The negative relation between due diligence and deal activity is even stronger if we take first differences and fix the timing. For a deal closed in quarter t, much of its diligence likely takes place in quarter $t-1$. Therefore, we next relate diligence levels in quarter $t-1$ to deal volume in quarter t. Table 3 shows that shifting the timing in this way changes the correlation from -33% to -45%, where the latter is significant at the 5% level. If we further take first differences to sweep out slow-moving confounding variables, the correlation changes to -66%, significant at the 1% level. Figure 5 plots that first-difference relation. Quarter 2 of 2021 stands out for its large increase in deal volume and large decrease in diligence, while quarter 3 of 2022 stands out for the reverse. One concern with this aggregate time-series analysis is that in-person meetings were disrupted by the COVID-19 pandemic. When we use only quarters after 2021q1, when in-person meetings largely resumed, the correlation strengthens further, to -84%. These correlations are provocative, but they face obvious limitations: the small observation count makes inference difficult, and confounding aggregate shocks make interpretation difficult.

Our main tests related to the due diligence choice are in Table 4. This table shows results from deal-level regressions with dependent variable equal to the log of our diligence measure. We include only observations with a strictly positive diligence measure. Across the columns we include an increasing number of fixed effects (FEs). Month FEs soak up aggregate shocks, including the rise of virtual meetings and other effects of COVID-19, which altered diligence patterns during our sample period. Month FEs also absorb how hot or cold the overall VC industry is during each period. Industry and stage FEs control for unobserved costs and benefits of due diligence at these levels.[11] For example, if earlier-stage companies face higher uncertainty and therefore require more diligence, or if in-person diligence is relatively more

[10] According to TDK Ventures (2024), "Throughout 2021, dealmaking—and due diligence—experienced historic changes. [...] FOMO (fear of missing out) was rampant. [...] The acceleration in dealmaking came with compressed deal cycles and less robust vetting. Power dynamics shifted to founders, and there was a growing sense that rigorous, time-consuming diligence was not 'founder-friendly.' To get in on hot deals and look good to their LPs (limited partners), VCs began prioritizing speed and cutting corners on diligence. [...] There was often an assumption among VCs that someone else had already done the diligence, especially for startups that had a lot of hype surrounding them. Preemption became the norm as VCs rushed to get founders term sheets ahead of a round. This frenetic pace of investing was embodied by crossover hedge fund Tiger Global. Tiger Global closed 354 VC deals in 2021, nearly one per day. In some cases, it would get to a term sheet in as little as 48 hours." By 2022, the mood had changed dramatically: "The slower market has given investors time to once again conduct due diligence [...]" (Davis and Miller, 2022).

[11] We use the seven industry categories from Table 2. Stage takes on values Seed, Series A, Series B, and so on. See Table 4 for details.

important at younger companies, those effects would be absorbed by the stage FEs. Column 5 shows our most saturated model, which compares deals within the same stage and same industry-by-month.

Panel A shows our baseline tests, which use two proxies for the costs of due diligence. The first regressor, log(Distance), is the log geographic distance between the investor's and startup's office buildings. This variable proxies for travel costs, a direct cost of due diligence. This regressor's estimated coefficient is negative and statistically significant at the 1% level in all columns. Its value remains fairly stable as more FEs are introduced. The coefficient in the first column implies that a doubling of distance is associated with a 35% reduction in diligence hours.[12]

Panel A's second proxy for diligence costs is log(VC Contacts per Month), the log number of other VC firms meeting with the focal startup per month leading up to the deal. When calculating the number of other VCs, we exclude the focal VC, all co-investors in the current round, and all investors from previous rounds. This variable proxies for how hot the deal is. The intuition behind this proxy is that, when a startup is contacted by more investors per month, there is a higher likelihood that the deal is taken away by another investor during the diligence period, thus increasing the indirect cost of conducting more due diligence. The proxy's coefficient estimates are negative and statistically significant at the 5% level across all columns. These negative coefficients are consistent with VCs doing less diligence when facing a higher risk of losing the deal. The coefficient estimates decline only slightly in magnitude as more FEs are included. The coefficient in the last column implies that a doubling of VC contacts per month is associated with a 13% reduction in diligence hours.

VC Contacts per Month has the virtue of measuring the level of competition specific to each deal. Panel B studies an alternative measure of competition constructed at the year-stage-industry level. Abnormal Deal Volume equals the deal volume for the same year, stage, and industry as the focal deal, divided by the average deal volume over the previous two years. A higher value indicates that the startup's sector has gotten hotter in the given year. Since different sectors of the VC market go through booms at different times, we can include time FEs to soak up aggregate trends and gain identification from variation across sectors. Similar to Panel A, we find a significantly negative relation between the amount of due diligence and how hot the sector is. The relation is significant only at the 10% level in columns 2 and 5. Even with month, industry, and stage FEs in column 4, the coefficient remains negative and statistically significant at the 5% level. Economic significance is quite

[12]The estimated coefficient is -0.632, so doubling Distance changes the diligence measure by a fraction $2^{-0.632} - 1 = -0.35$.

high. Taking the coefficient smallest in magnitude, from column 1, we find that a one standard deviation increase in abnormal deal volume is associated with a 15% reduction in diligence.[13] The magnitude increases to a 34% reduction if we use the largest coefficient, from column 5.

In Panel C, we replace geographic distance with a different proxy for the direct costs of due diligence. Log(Deals per Partner) is the log ratio of (1) number of deals the investment firm participates in within 18 months of the focal deal's investment date to (2) the number of investment professionals at the investment firm. The latter count, from PitchBook, includes roles such as principals, partners, directors, and associates, but excludes positions like admins and accountants. We predict that as the investors become busier with more deals, the opportunity cost of their time increases, which leads them to perform less diligence. The negative coefficients on log(Deals per Partner) in Panel C support that story. Even in the presence of stage and industry-month FEs, the coefficient is significant at the 1% level and implies a 22% decrease in the diligence level in association with a doubling of deals per partner.[14]

How does measurement error in our diligence proxy affect these results? First, we measure in-person meetings imperfectly. If this measurement error is random (i.e., classical), and since it affects the dependent variable, it does not bias our regression results. Second, our measure omits other types of diligence besides in-person meetings, like virtual meetings. If we are only interested in how in-person meetings relate to our regressors, this omission does not matter. But if we care about total, unobservable diligence, then the omission can bias our results. Let β denote the coefficient of total diligence on our regressors. We consider three cases: (1) If the omitted diligence is unrelated to the regressors, then our regressions estimate β without bias. (2) If a regressor causes in-person and omitted diligence to move in opposite directions, then our regressions overstate β's magnitude. For example, VCs can substitute virtual for in-person meetings when distance is greater (Alekseeva et al., 2025), making the relation between distance and total diligence less negative than our estimated coefficient would suggest. (3) If a regressor moves both types of diligence in the same direction, then our regressions can understate β's magnitude. For instance, if VCs cut all forms of diligence in "hot deals," total diligence decreases more than just in-person meetings do; hence, our estimated coefficient understates how much "hot deals" reduce total diligence.[15]

[13] The coefficient estimate is -0.412, and the standard deviation of abnormal deal volume is 0.394, so $\exp(-0.412 \times 0.394) - 1 = -15\%$.

[14] Raising a new fund can also make VCs busy. However, we find no significant relation between diligence levels and whether the VC firm raises a new fund within the next two years.

[15] The logic is somewhat complicated by the fact that we take logs of the diligence measure. If the VC cuts all forms of diligence by the same factor, then β would accurately capture the log change in total diligence.

To summarize, we find that longer due diligence is negatively related to its costs. We study proxies for direct costs, such as greater distance and busier investors. We also study whether the deal, its sector, or the overall VC market is hot, arguing that due diligence has higher indirect costs when VCs face more competition, more investment opportunities, or less bargaining power relative to startups.[16] Since we do not have exogenous variation in these costs, these results do not have a causal interpretation. The results are consistent, however, with a simple model of the due diligence choice. We present that model next.

5. A model of due diligence and capital allocation

We provide a simple theory to help interpret the previous results on the due diligence choice, and also to generate predictions about capital allocation and investment outcomes. We test the latter predictions in the next section.

We model due diligence as producing a signal about the quality of the startup-VC match. The investor chooses how precise a signal to obtain, analogous to how much due diligence to perform. This choice involves a tradeoff: learning is costly but allows a more profitable investment choice. By modeling investor learning as a choice, our model is similar in spirit to Grossman and Stiglitz (1980) and, more recently, Daley et al. (2024).

5.1. Model setup

The model features a single VC, a single startup, and two periods. The VC maximizes the expected surplus from investing in the startup:

$$\max_{\tau, K} E\left[aK^{\theta} - K - c\tau\right]. \tag{1}$$

The first term, aK^{θ}, is the startup's valuation step-up from this financing round to the next, i.e., the change in the startup's value from before the VC's arrival to just before the next financing round occurs. The step-up depends on the interaction between K, the amount of new money invested in the startup, and a, the unobservable quality of the VC-startup match. That match quality reflects both the startup's quality and the VC's ability to add value. The curvature parameter θ is in the interval (0,1). The second term, $-K$, reflects the investment's direct cost. The third term, $-c\tau$, reflects the direct and indirect costs of due diligence. The VC chooses τ, the precision of the signal obtained through due diligence. We

[16] Wangerin (2019) finds a similar phenomenon in mergers and acquisitions: competitive pressures are associated with shorter due diligence.

interpret choosing a higher τ as choosing to perform more due diligence. Parameter c is the cost per unit of precision. We interpret c as the cost per unit (e.g., hour) of diligence.

At $t = 0$, the VC's beliefs about a are distributed as

$$\log(a) \sim N(\mu_0, 1/\nu_0), \tag{2}$$

where μ_0 and ν_0 are the prior's mean and precision, respectively. The VC chooses τ at $t = 0$. Due diligence occurs at $t = 1$ and produces a signal S distributed as

$$S \sim N(\log(a), 1/\tau). \tag{3}$$

The VC chooses K immediately after observing S. At $t = 2$, interpreted as the time of the next financing round or exit, the value of a is realized.

5.2. Solution and testable predictions

The Appendix shows the full model solution and proofs. We start by analyzing the problem at $t = 1$. Standard results on Bayesian learning deliver the VC's beliefs after seeing the signal:

$$\begin{aligned} \log(a)|S &\sim N(\mu_1, 1/\nu_1), & (4)\\ \mu_1 &= \left(\frac{\nu_0}{\nu_0+\tau}\right)\mu_0 + \left(\frac{\tau}{\nu_0+\tau}\right)S, & (5)\\ \nu_1 &= \nu_0 + \tau. & (6) \end{aligned}$$

After seeing S, the VC chooses K by taking the first-order conditions of equation (1), treating the choice of τ from the initial period as given. The optimal choice of K is

$$K^* = (\hat{a}\theta)^{\frac{1}{1-\theta}}, \tag{7}$$

where $\hat{a} = E[a|S]$ is expected match quality after observing the signal. A higher signal S leads to higher perceived quality $\hat{a}$, which in turn leads the VC to invest more in the startup.

At $t = 0$, the investor chooses τ by trading off the costs and benefits of learning. The costs of learning are from parameter c. Intuitively, if learning is more costly, the VC chooses to learn less. We formalize this prediction next.

Prediction 1. *The chosen τ decreases in c. A sufficient condition for this result to hold is that*

$$\tau > \frac{\theta}{4(\theta-1)^2} - \nu_0. \tag{8}$$

This prediction supports the previous section's empirical results showing a negative relation between the costs and chosen levels of due diligence.

To understand the benefits of learning, it helps to rewrite the objective function as

$$\max_{\tau} E[a(K^*)^\theta - K^* - c\tau], \tag{9}$$

recalling that the chosen K^* is a function of $\hat{a}$, which is a function of the signal S. If signals are free (i.e., $c = 0$), and treating the true a as given, the objective function above is maximized at $\hat{a} = a$. This means the investor would choose to learn the match quality perfectly, i.e., $\tau = \infty$. The benefit of learning is that it allows the investor to make a more profitable choice of K^*.

To study capital misallocation, we follow Hsieh and Klenow (2009) and the large literature that follows it. These papers typically use the variance of MPK as a measure of capital misallocation. We take a similar approach here. In our setting, the MPK of a given deal is

$$MPK = \frac{d}{dK}(aK^\theta). \tag{10}$$

Evaluating the MPK at the chosen capital level, K^*, yields

$$MPK = a\theta(K^*)^{\theta-1} = a/\hat{a}, \tag{11}$$

where the second equality uses equation (7). If the VC receives a signal that is "too high" (i.e., above a), then they overestimate match quality (i.e., $\hat{a} > a$), leading the VC to over-invest. As a result, MPK < 1. If instead the signal is "too low," then the investor underestimates match quality ($\hat{a} < a$), leading to under-investment and MPK > 1. In either case, capital is misallocated relative to the perfect-information benchmark. In that benchmark, $\hat{a} = a$ and the MPK equals 1, with no dispersion. In general, more learning results in less dispersion of MPK, as formalized below.

Prediction 2. *$Var(\log(MPK)) = 1/(\nu_0 + \tau)$, which decreases in τ.*

Intuitively, the less the VC learns, the farther the VC's beliefs are from the truth, the more over- and under-investment occurs, and the greater the dispersion in MPK. As in Hsieh and Klenow (2009), the variance of MPK measures capital misallocation in an ex-post sense. In our setting, this misallocation results from imperfect information.[17] The formula above shows that dispersion in MPK also decreases with ν_0, the precision of prior beliefs about the VC-startup match. This result is also intuitive: deals with more prior uncertainty have

[17]Similar to Prediction 2, the model of David et al. (2016) implies that the variance of log MPK equals the variance of posterior beliefs about log productivity; see their equation (9).

more volatile outcomes. One striking feature of Prediction 2 is that the predicted variance of MPK depends only on τ and ν_0, so there is no predicted role for the chosen investment amount K^*, curvature θ, or cost c.

Interestingly, the model predicts no relation between τ and the average *level* of MPK.

Prediction 3. $E[MPK] = 1$ *for any value of* τ*.*

The intuition here is that VCs are rational Bayesian learners whose beliefs are correct on average, regardless of how much they choose to learn. While they sometimes over- or under-invest, causing MPK to diverge from 1, on average they get it right, producing average MPK=1 even when they choose to learn very little.

According to the model, extra due diligence creates value (via better capital allocation) and yet does not increase expected performance (as measured by MPK). To resolve this tension, we must distinguish between dollar and percent returns. Over- and under-investing both reduce the total dollars of value created, but they have opposite effects on value created per dollar invested, i.e., MPK. Under-investing delivers a high MPK, but the VC would have created more total dollar value by investing more, even though MPK would be lower.

In practice, valuation is at the center of negotiations between VCs and startups. Our model can be interpreted in that light. In the Online Appendix, we extend the model to assume that the VC and startup bargain with each other to split the deal's surplus, which amounts to bargaining over the deal's pre-money valuation. In place of equation (1), we assume the VC maximizes its expected dollar profit from the deal. We show that the baseline and extended models make the same predictions about K and τ, regardless of the VC's bargaining power.[18] The extension also delivers the intuitive result that, for a given investment amount K, VCs with more bargaining power obtain larger fraction ownership stakes, meaning they invest at lower pre-money valuations.

6. Evidence on capital allocation and performance

This section empirically studies the implications of due diligence for capital allocation and investment performance. Section 6.1 discusses measurement. Section 6.2 shows tests related to the variance and level of MPK (Predictions 2 and 3). We also relate due diligence levels to VC returns. Section 6.3 interprets the reduced-form evidence through the lens of our model, with the goal of quantifying the value gained from due diligence.

[18] The extension requires a few additional assumptions, such as an adjustment to the cost parameter c.

6.1. Measuring MPK, returns, and their variance

Testing Predictions 2 and 3 requires a proxy for MPK. Fortunately, our model provides some guidance. We rewrite equation (11) as

$$MPK = \theta \frac{a(K^*)^\theta}{K^*}. \tag{12}$$

The fraction's numerator is the financing round's valuation step-up, according to equation (1). Therefore, we measure MPK as the round's valuation step-up divided by the amount invested in the round (the denominator). This measure equals MPK up to the proportional constant θ, which we assume is equal across observations.[19] A virtue of this measure is that it accounts for curvature in the production function without requiring a direct estimate of θ. Another virtue is that it does not require an estimate of match quality a, which our model assumes is unobservable at the time of the investment but is realized by the startup's subsequent financing round.

We measure the amount invested in a given financing round using the PitchBook variable 'Deal Size.' We measure the valuation step-up as $PreMV_{next}/(1+r) - PreMV_{current}$. $PreMV$ refers to a financing round's pre-money valuation, a measure of the startup's market price implied by the financing round, measured immediately prior to the round's injection of new capital. $PreMV_{next}$ refers to the pre-money valuation in the startup's subsequent financing round, and $PreMV_{current}$ is the pre-money valuation in the current round. To compute pre-money valuation, we divide the PitchBook variable 'Deal Size' by 'Percent Acquired' to obtain the post-money valuation, then subtract 'Deal Size' to get the pre-money valuation.[20] Next, we discuss three measurement challenges and explain the role of r.

First, valuation step-ups in the data are affected by unexpected shocks that hit the startup between the two financing rounds. Those ex-post shocks, which are outside our model, add measurement error to both the estimated level and variance of MPK. The shocks also bias our estimated $Var(\log(MPK))$ upward. Measurement error will not necessarily bias our regression coefficient estimates, since the MPK measures are on the left-hand side of our regressions, and it is unlikely that the shocks' volatility is correlated with our due diligence measures. The upward bias in $Var(\log(MPK))$ is also not necessarily a problem, as we are not trying to measure the absolute level of this variance. Instead, our goal is to

[19] Even if θ varies across deals, our regression coefficients are not necessarily biased. We take logs of our MPK proxy, so our regression residuals include $\log(\theta)$ or its variance. As long as those objects are not related to our due diligence measure, then the coefficients of interest are not biased.

[20] If there is an exit instead of a subsequent financing round, we replace $PreMV_{next}$ with the estimated exit value where possible. For instance, if the exit is an M&A, we calculate the exit value by dividing the PitchBook variable 'Deal Size' by 'Percent Acquired.'

relate $Var(\log(MPK))$ to our due diligence measures. As long as the shocks' volatility is unrelated to our due diligence measures, then we have not introduced bias into our test.

Measurement error nevertheless can reduce the power of our tests. To address that concern, we soak up some of the ex-post shocks in three ways. (1) To soak up industry-level shocks that hit the startup between rounds, we divide $PreMV_{next}$ by $1 + r$, where r is the return on the industry stock portfolio matching the startup's industry.[21] We measure that return between the dates of the startup's current and next financing rounds. By dividing $PreMV_{next}$ by $1+r$, we measure the hypothetical valuation step-up assuming there had been a zero return on the startup's sector. (2) When our dependent variable is $Var(\log(MPK))$, we control for the startup's age and duration between the two financing rounds; both controls can relate to the volatility of shocks hitting startups between rounds. (3) We continue including many fixed effects in our regression. Stage FEs control for the fact that shocks are more volatile at some stages than others. Month FEs soak up aggregate shocks, and industry-by-month FEs further soak up sector-level shocks.

The second measurement challenge is that PitchBook is often missing the data required to compute pre-money valuations, and successful startups are more likely to disclose their valuations (Cochrane, 2005; Korteweg and Sørensen, 2010). This selection biases upward the average level of valuation step-ups, but that level is not the object we study. Instead, we are interested in the relation between due diligence and investment outcomes. Below, we analyze how selection bias may affect these relations. Also, for robustness we show that imputing zeros for some of the missing valuations does not change our conclusions (see Section 8).

Third, complex VC deal structures can make pre-money valuations poor approximations of companies' market prices (Gornall and Strebulaev, 2020; Metrick and Yasuda, 2021). This is yet another source of measurement error. Mitigating this concern, the measurement error in $Var(\log(MPK))$ is on our regressions' left-hand side, and the errors are plausibly unrelated to the due diligence measure we study.

We also proxy for VC returns, although our model does not offer predictions about them. Lacking VC cash-flow data that can be merged to our sample, we approximate the VC's return as the log change in the value of the VC's stake from this round to the next. Specifically, our return measure is the log ratio of the next round's pre-money valuation to the current round's post-money valuation. Lining up the pre- and post-money valuations in this way produces a measure unaffected by the amount of capital injected in the current

[21]We map the startup's 'Primary Industry Sector,' from PitchBook, to an S&P 1500 industry index. For example, we map PitchBook's 'Information Technology' to 'SP1500 Information Tech.S.' Details are in Appendix Table A.1.

round. Of course, this measure faces the same challenges discussed above.

Testing Prediction 2 requires measuring not just MPK but also its variance. We model variance starting with the definition $Var(Y|X) = E[(Y - E[Y|X])^2|X]$. Setting Y equal to $\log(MPK)$ and X to the set of regressors, we first regress Y on X. We store the regression's fitted values, which correspond to $E[Y|X]$, as $\overline{\log(MPK)}$. Our proxy for $Var(\log(MPK_i))$ for deal i is then the squared value of $(\log(MPK_i) - \overline{\log(MPK_i)})$, which corresponds to $(Y_i - E[Y_i|X_i])^2$. By the definition above, this variable's expected value is $Var(\log(MPK_i)|X_i)$.[22] Since we have the variable's realized rather than expected value, we measure variance with error, but this error is in our dependent variable and therefore does not introduce bias. We model the variance of returns similarly.

6.2. Reduced-form evidence

Table 5 shows our tests of Prediction 2. Our theory predicts a negative relation between the variance of MPK and the level of due diligence. We estimate deal-level regressions with dependent variable equal to the previously defined proxy for the variance of $\log(MPK)$. The regressor of interest is the log of our due diligence measure. We use the log transformation to tame outliers in our measure and approximate the nonlinear relation from Prediction 2.

Panel A shows results from simple specifications that include fixed effects but no other controls. Consistent with Prediction 2, dispersion in MPK decreases in the amount of due diligence. The relation is statistically significant at the 5% level in all specifications except the last, which includes the most granular fixed effects. Even there, the coefficient remains negative.

To judge economic significance, we take the coefficient on $\log(DD)$ in column 4, the last column with statistical significance. There, we find that a one standard deviation increase in $\log(DD)$ is associated with a 0.12 decline in the variance of $\log(MPK)$.[23] If we start from the sample average volatility of $\log(MPK)$, 130%, increasing $\log(DD)$ by one standard deviation reduces the volatility of $\log(MPK)$ to 125%.[24] This decline appears modest primarily because ex-post shocks make the measured dispersion in MPK very high. If we could instead

[22] Alternatively, we could model variance through maximum likelihood estimation, and within that estimation we could estimate how variance relates to our due diligence measures. We suspect results would be very similar. Our current approach has the benefit of being simpler and more transparent.

[23] The standard deviation of $\log(DD)$ is 3.5, and the coefficient estimate is -0.034, so $-0.12 = 3.5 \times (-0.034)$.

[24] The sample mean $Var(\log(MPK)) = 1.69$, corresponding to a standard deviation of $\sqrt{1.69} = 130\%$. Reducing the variance from 1.69 to $1.69 - 0.12$ changes the standard deviation to 125%.

measure the perceived dispersion in MPK at the time of the VC's investment—the dispersion to which our model refers—then the effect size would appear relatively larger. Also, measurement error in $\log(DD)$ biases the coefficient estimates toward zero, so we underestimate the true magnitudes. Section 6.3 explores economic significance in more depth.

According to Prediction 2, the dispersion in MPK depends not just on the amount of due diligence but also on prior uncertainty. Prior uncertainty is a potentially important omitted variable, as it also influences the choice of due diligence intensity. Prior uncertainty surely varies across stages, industries, and years. The fixed effects in column 4 absorb those effects, partially controlling for prior uncertainty. The fixed effects also control for variation across time, industries, and stages in the volatility of ex-post shocks. We see that these FEs, especially the stage FEs, soak up a good deal of the regression's residual volatility.

To more carefully control for prior uncertainty and features omitted from the model, we add control variables in Panel B of Table 5. We find that adding these controls typically increases the estimated magnitudes of the due diligence coefficients, but only slightly. The controls do soak up residual variance, which increases the due diligence coefficients' statistical significance. We control for the log of startup age as a proxy for prior uncertainty, arguing that younger startups face more uncertainty. Consistent with Prediction 2, startup age enters with a negative coefficient. To further control for prior uncertainty, we include a measure of the VC's relevant industry experience, measured as the proportion of the VC's recent deals (within 18 months) in the same industry sector as the focal deal. A VC with more industry-specific experience arguably perceives a startup more clearly even before due diligence begins. Consistent with this logic, we see a significantly negative relation between industry experience and MPK dispersion, further supporting Prediction 2.[25] Although Prediction 2 does not predict a role for K (the amount invested in the round), we include it as a control and find it has a strongly negative relation with dispersion in MPK. A potential story outside our model is that VCs' risk aversion leads them to allocate less money to deals facing more uncertainty. In addition, we control for the duration between this round and the next in order to soak up variation in ex-post shock volatility, which is omitted from our model. If more time elapses between this round and the next, there is more time for shocks to hit the startup, so we anticipate more ex-post dispersion in valuation step-ups. Indeed, this control has a strongly positive coefficient.

Panels A and B use the full sample, including the many deals where we record zero due

[25] Alternatively, a VC with more industry experience can learn more per hour of due diligence. The same regression prediction obtains: for a given number of hours of due diligence, a VC with more industry experience learns more, reducing their posterior variance of beliefs, which reduces the variance in MPK.

diligence. To include those observations while still taking logs, we use the transformation $\log(1+X)$.[26] The case for including those zeros is that, while noisy, they contain information: a zero indicates a higher probability that the true, total, unobservable amount of due diligence is very low. The case for dropping the zeros is that, relative to non-zeros, they are more likely to reflect measurement error. As a robustness check, in Panel C we repeat the baseline regressions from Panel A using only observations with a positive due diligence measure. We are left with less than 250 deals.[27] Despite that small sample, we continue to find a significantly negative relation between diligence intensity and dispersion in $\log(MPK)$, supporting Prediction 2. The relation is significant at the 1% level in three specifications and insignificant in only one. Economic significance is even larger than before. Again studying column 4, a one standard deviation increase in log diligence is associated with a -0.49 decrease in variance of MPK, whereas in Panel A the magnitude is -0.12.

To summarize, in almost all specifications we find a significantly negative relation between the level of diligence and the dispersion in MPK. Interpreted through our model, the result indicates that less due diligence leads VCs to invest under greater uncertainty, which produces more extreme investment outcomes. Those extremes reflect more over- or under-investment—that is, more capital misallocation.

How does measurement error in our due diligence proxy affect these results? Dispersion in MPK arguably depends on total due diligence, not just the in-person meetings we observe. Our goal is to estimate the coefficient, denoted β here, of MPK dispersion on total due diligence. Any unmeasured forms of due diligence end up in the regression's error term. If these unmeasured parts are uncorrelated with the in-person meetings we do measure, our estimates of β are unbiased. However, if measured and unmeasured diligence are substitutes—such that they are negatively correlated—then the omission attenuates our estimates of β toward zero. Another concern is classical measurement error in our proxy for in-person meetings, which further attenuates the estimate of β. This attenuation biases us away from finding support for the model's prediction about due diligence and MPK dispersion.

Since growing startups with high valuation step-ups are more likely to disclose, the observed data may exclude many low-MPK startups. In the Online Appendix, we use our model to analyze any resulting selection bias. We show that selection would bias the coeffi-

[26]Chen and Roth (2024) show that the $\log(1+Y)$ transformation for outcome variables Y makes it hard to interpret economic significance. When judging economic significance above, we take the $\log(1+X)$ transformation into account.

[27]In the previous section, where we study the due diligence choice, we focus only on the sample with a positive diligence measure. There, we do not need data on valuation step-ups, so we have a much larger sample. In this section, requiring both a non-missing valuation step-up and a positive diligence measure produces a small sample.

cients on due diligence in Table 5 upward, toward zero. This bias also pushes in the opposite direction of our main findings in Table 5.

Interpreting dispersion in MPK as evidence of misallocation requires caution, as Hsieh and Klenow's (2009) approach has well-known limitations. First, measurement error in MPK can inflate its dispersion, leading to overestimation of misallocation (e.g., Bils et al., 2021). Mitigating this concern, we focus not on the unconditional level of MPK dispersion but instead on its relation to due diligence. As long as the variance of measurement error in MPK is unrelated to our diligence proxy, which seems plausible, this bias does not affect us. Second, the Hsieh-Klenow approach relies on a Cobb-Douglas functional form, which may not well approximate reality. Another well-known concern is that measuring MPK with revenue data contaminates MPK with markups. We avoid this problem by using data on valuation step-ups rather than revenues. Finally, real frictions can produce dispersion in MPK, but we should not necessarily interpret such dispersion as misallocation.

In both our model and the regressions, the level of due diligence is endogenous. Rather than estimating causal relations, we estimate relations between equilibrium quantities. We have done our best to control for prior uncertainty, a potential omitted variable, but some variation in prior uncertainty may remain. For example, there can be more prior uncertainty about startups with more soft information. The resulting omitted-variable bias, however, works in the opposite direction of our results. In our model, higher prior uncertainty produces more volatile outcomes and leads the investor to perform more diligence ex ante, but empirically we find a negative relation between the two.

Next, we study the level of MPK and test Prediction 3. Similar to the previous analysis, we work at the deal level and regress the log level of MPK on the amount of diligence. Results are in Table 6. Consistent with Prediction 3, we find no statistically significant relation between the level of MPK and the amount of due diligence, at least in Panel A. When we add controls in Panel B, the relation turns significantly negative in the simplest specifications but loses significance when we include stage FEs. In Panel C, we again study the sample with only positive diligence, and we find a significantly negative relation in all but the simplest specification. The evidence is therefore mixed. Some results support our model's predicted non-relation. Other results point to a negative relation, suggesting that less due diligence produces more under-investment and hence higher MPK. Cautioning against the negative relation, we show in Appendix B.2 that a negative relation can arise spuriously if less-successful startups are less willing to report the data we need to compute MPK.

The model has clear implications for MPK but not for VCs' returns. Returns are of independent interest, however, so we study them next. Our return measure has a 0.87 correlation with MPK, in logs. These two concepts are related but distinct. MPK captures how injecting K in new capital changes the startup's total value; that change includes both the direct effect of injecting cash into the balance sheet and the indirect effect of the startup and VC working together to invest the cash in good projects. Our return proxy captures the fraction change in the company's value, and hence the fraction change in the value of VC's stake, from immediately after the financing round until the next financing round.

Tables 7 and 8 mimic Tables 5 and 6, except we replace MPK with our proxy for the VC's return. Results are qualitatively similar to before. Table 7 shows a consistently negative relation between the due diligence measure and the variance of the VC's return. The result loses significance only in Panel C, where we work with the small, positive-diligence subsample. Even there, the relation is highly significant in the most saturated specification. As before, economic significance appears modest mainly because ex-post shocks make the level of return variance very high.[28] Similar to before, Table 8 shows a negative but typically insignificant relation between the due diligence measure and the level of return.

Finally, we return to the aggregate time-series patterns discussed earlier. If there is less due diligence during hot markets, and if less due diligence produces more misallocation, then we should see more evidence of misallocation in hot markets.[29] Consistent with that prediction, we find a 31% time-series correlation between aggregate VC deal volume and MPK dispersion. The correlation increases to 58% and becomes statistically significant at the 1% level if we work in first differences. See Figures B.4 and B.5.

6.3. Quantitative implications

To what extent does due diligence improve the allocation of venture capital? We perform a simple quantification to address this question. By combining the reduced-form evidence from the previous subsection with the model from Section 5, we estimate how a hypothetical increase in due diligence would increase the value created by a VC investment.

We begin by defining the value created by the VC financing round as

$$\Pi(a, K) = aK^{\theta} - K, \tag{13}$$

[28]For example, if we take the median slope coefficient in Panel A, we find a one standard deviation increase in the diligence measure is associated with a -0.025 decrease in return variance. To put that magnitude into context, it would reduce return volatility (the square root of variance) from its average of 91% to 89%.

[29]This pattern aligns with Aran and Packin (2024), who argue that cyclical market conditions exacerbate the "due diligence dilemma," in which VCs sacrifice scrutiny for speed in hot markets.

which equals the round's valuation step-up less the round's financial cost. Recall from equation (7) that, given the investor's beliefs $E[a] = \hat{a}$, the investor chooses

$$K^* = (\hat{a}\theta)^{\frac{1}{1-\theta}}. \tag{14}$$

If the investor could instead observe the true match quality, a, they would choose

$$K^{**} = (a\theta)^{\frac{1}{1-\theta}}. \tag{15}$$

The effect of imperfect information (i.e., seeing $\hat{a}$ instead of a) on value creation is

$$\Delta\Pi = \Pi(a, K^*) - \Pi(a, K^{**}). \tag{16}$$

The first argument in both terms must be the true a, because a is the actual match quality affecting the deal's value creation, regardless of the investor's beliefs and chosen K.

In the Appendix, we show that $\Delta\Pi$, as a fraction of the amount invested, is a function of the deal's MPK:

$$\frac{\Delta\Pi}{K^*} = \frac{1}{\theta}\left(MPK - MPK^{\frac{1}{1-\theta}}\right) - \left(1 - MPK^{\frac{1}{1-\theta}}\right) \leq 0, \tag{17}$$

recalling that $MPK = a/\hat{a}$. Further, $\Delta\Pi/K^*$ reaches a maximum of zero when $MPK = 1$, which is when the beliefs equal the truth: $\hat{a} = a$. When beliefs diverge from the truth, $\Delta\Pi/K^*$ is strictly negative, meaning imperfect information destroys value. Taking expectations of the equation above yields

$$E\left(\frac{\Delta\Pi}{K^*}\right) = \frac{1}{\theta}\exp\left(\mu + \frac{\sigma^2}{2}\right) + \left(\frac{\theta - 1}{\theta}\right)\exp\left(\frac{\mu}{1-\theta} + \frac{\sigma^2}{2(1-\theta)^2}\right) - 1, \tag{18}$$

where μ and σ^2 are the mean and variance of $\log(MPK)$. That expression equals the average value lost, expressed as a fraction of capital invested, due to imperfect information.

To help clarify this concept, consider the following simple example. Suppose we had the opportunity to invest in a startup. If we knew the true match quality, a, the optimal investment would be \$4M, leading to a valuation step-up of \$6M and hence value creation of \$2M (i.e., \$6M - \$4M). However, in reality, we cannot observe the true quality, so we conduct some due diligence, form a belief $\hat{a}$, and suboptimally decide to invest \$2M. This results in a valuation step-up of \$3M and value creation of \$1M (i.e., \$3M - \$2M). Since imperfect information reduced the value created from \$2M to \$1M, the difference of \$1M is referred to as the value lost from imperfect information. This value lost, as a fraction of capital invested, is 50% (i.e., \$1M/\$2M=50%), which corresponds to $\Delta\Pi/K^* = -50\%$.

Using equation (18), we can connect the regression coefficients in Table 5 to the average value lost due to imperfect information. Doing so requires estimates of the mean and variance of $\log(MPK)$ (μ and σ^2, respectively) and the curvature parameter θ. The Appendix explains how we estimate these parameters using data on valuation step-ups and investment amounts, while accounting for ex-post shocks. The baseline estimates we use are $\hat{\mu} = -0.06$, $\hat{\sigma}^2 = 0.54$, and $\hat{\theta} = 0.21$. Given the challenges in estimating θ, we show results for a range of values centered around $\hat{\theta}$. We substitute these parameter values into equation (18) to calculate a baseline value of the variable of interest, $E\left(\Delta\Pi/K^*\right)$. We quantify how this average value lost due to imperfect information would change if the hours spent on due diligence hypothetically doubled. To calculate this change, we keep θ and μ constant but adjust σ^2 by combining it with the coefficient of interest from Table 5. For example, a coefficient of -0.07 in Table 5 suggests that doubling due diligence hours (increasing $\log(DD)$ by $\log(2)$) would reduce $\sigma^2 = Var(\log(MPK))$ by 0.05 ($= -0.07 \times \log(2)$). We then use equation (18) to measure how $E\left(\Delta\Pi/K^*\right)$ changes if σ^2 changes from its baseline value of 0.54 to 0.49 ($= 0.54 - 0.05$), keeping θ and μ at their baseline values.[30]

Figure 6 shows the effect of doubling the amount of due diligence on the average value lost, considering different values of θ and various coefficients of interest from Table 5. Those coefficients are approximately -0.03, -0.05, and -0.07 in Table 5 Panels A and B, and they range from -0.10 to -0.20 in Panel C. We plot five curves corresponding to these values. As a baseline, we consider the -0.07 coefficient from Table 5 Column 1 Panel A, and we use the point estimate $\hat{\theta} = 0.21$, indicated with a vertical dashed line. In this baseline, the corresponding y-axis value is roughly 6%. This means that doubling due diligence hours would reduce the average value lost from imperfect information by 6% of the VC's amount invested.[31] Multiplying 6% by the \$166B of aggregate VC investment in the U.S. in 2023 (PitchBook, 2024) translates into an extra \$10B in aggregate value saved from a hypothetical doubling of VC due diligence. These magnitudes are substantial.

Are the magnitudes plausible? Consider a \$3M early-stage VC investment; this is the median deal size in 2023 (PitchBook, 2024). Assume the VC currently spends 118 hours on due diligence; this is the average value from the Gompers et al. (2020) survey. Doubling the hours spent on due diligence would deliver an extra \$180K (calculated as \$3M $\times$ 6%) of value in return for an extra 118 hours of work, producing \$1525 of extra value per hour of work. This number seems plausible, in the sense that the opportunity cost of a VC's

[30]Recall that 0.54 is the baseline estimate of σ^2 , calculated as the sample variance of the ratio of valuation step-up to capital invested, minus the estimated volatility of ex-post shocks. The detailed steps are provided in the Appendix.

[31]Going back to the simple example above, this estimate means that doubling due diligence increases the value created from \$1M to \$1.12M=\$1M + 6%$\times$2M.

time might exceed this amount, in which case the VC would choose to stay at 118 hours of diligence rather than double it. In other words, the extra effort on due diligence might not be privately optimal. Even if the VC's time were worth less than \$1525 per hour, spending extra time on due diligence could be suboptimal, because doing so could allow a competing, faster VC to steal the deal.

Our model omits many features of reality, so we do not interpret the calibrated magnitudes above literally. Also, Figure 6 shows the possible magnitudes cover a wide range. If we use the slope coefficient of -0.20 instead of -0.07, then the improvement in value creation changes from 6% to 17%, and it increases further if we use a higher value of θ. On the low end, the improvement in value creation is 2.5% if we instead use the smallest coefficient magnitude along with $\theta = 0.15$. All these numbers are substantial, however. We simply conclude from this analysis that due diligence seems to play a large role in the efficient allocation of capital to startups.

Are observed levels of due diligence socially optimal? We would need a better model to answer that question. However, our results do suggest one reason why diligence levels may be below their social optimum. Recall that hot markets and deals are associated with less due diligence, potentially because VCs fear losing the deal to a competing investor. These competitive pressures would be irrelevant to a social planner. The planner would presumably conduct more due diligence in these scenarios, allowing more efficient capital allocation. In reality, coordination failures and competition among investors can lead to too little due diligence, from a social point of view.

7. Extensive margin and diligence on rejected deals

So far, we have focused on the intensive margin: variation in due diligence among completed deals. Yet VCs also conduct diligence on many deals that are ultimately rejected. As a final extension, we examine the extensive margin by studying both accepted and rejected deals. We find that diligence patterns across the full set of deals predict startup quality.

A deal can be rejected by either the VC or the startup. To focus on VCs' choices, we restrict attention to top VCs, which founders are less likely to turn down (Hsu, 2004). We define top VCs as those in the top quintile of assets under management, following Berk and Green's (2004) argument that higher-ability investors attract more capital. Our main proxy for startup quality is an indicator for whether the startup exits or raises a subsequent financing round. The analysis includes time, industry, and stage fixed effects to control for unobservables and end-of-sample effects. Section 8 explores robustness to these choices.

We find that startup quality increases monotonically across five groups of deals (Table 9). The first three groups consist of deals not led by a top VC. Within this set, startup quality rises with the extent of due diligence by top VCs. When no diligence is performed, average quality is 0.318, meaning 31.8% of startups successfully exit or raise a next round. That figure increases by 0.020 ($p = 0.004$) with limited diligence, and rises by another 0.028 ($p = 0.033$) with extensive diligence. Intuitively, conditional on being rejected, it is a positive signal when top VCs invest more time in evaluating the deal—much like a PhD candidate not hired by a top department but invited for fly-outs at those departments.

Comparing Groups 3 and 4, startup quality increases sharply, by 0.089 ($p = 0.013$), when deals move from being rejected after long diligence to being led by a top VC. This finding supports the view that top VCs have the skill to both identify and win high-quality deals. From Group 4 to Group 5, quality rises by another 0.046 ($p = 0.167$) when diligence changes from long to short, conditional on a top VC leading the deal. While our static learning model cannot explain this result, a dynamic learning model likely could. In such a setting, VCs optimally choose when to end their due diligence. Very positive early signals may justify stopping quickly to avoid further costs, whereas marginal deals may require extended evaluation before being accepted.

Previous research has already shown that funding by a top VC predicts startup success (e.g., Sørensen, 2007; Ewens and Rhodes-Kropf, 2015). We extend this literature by showing that diligence patterns themselves provide extra predictive power. Our findings suggest that matching on quality between VCs and startups begins as early as the diligence stage, and that top VCs deploy their diligence efforts strategically, choosing when to terminate the process based on the strength of observed signals.

8. Robustness

This section presents several robustness checks on our main empirical results from Section 6 and our extension in Section 7. Detailed results are in the Online Appendix.

First, we create an alternative diligence proxy equal to the time elapsed between the first VC-startup meeting and the deal date. While our main proxy captures active learning from in-person meetings, this alternative proxy captures other forms of diligence as well as passive learning, i.e., allowing information to emerge over time. Table B.1 shows counterparts of our main results from Tables 4 and 5. Eleven of the 14 originally significant coefficients remain statistically significant. Economic significance increases in some cases and decreases in others.

Our due diligence proxy includes meetings only with the round's lead investor. For robustness, we add meetings with co-investors in the round. Results are in Table B.2. The results on the due diligence choice weaken economically but strengthen statistically (Panel A), and vice-versa for results on MPK dispersion (Panel B). Some weakening of results is expected, as co-investors can spend time learning the same information, leading this alternative measure to double-count learning in some cases.

One limitation of our diligence proxy is that it relies on the address of the VC's headquarters, so we may miss meetings if the VC has multiple locations. Lacking data on all locations for each investor, we address the problem indirectly by noting that large investors are more likely to have multiple locations. We repeat our main analyses after dropping investors with AUM above the 90th or 80th percentiles (see Table B.3). We find that most results remain robust and some even strengthen.

Another concern is that geolocation data cannot distinguish between different floors of the same building. If the VC or startup is in a multi-tenant building, we might misclassify employees of other companies as VC or startup employees. Mitigating this concern, we measure only meetings between investors and startups in known, completed deals. It is unlikely that a different VC from Benchmark's office building, for example, would visit a startup funded by Benchmark. Nevertheless, for robustness, we repeat our main analyses after excluding observations where the VC or startup shares a building address with any of the roughly 360,000 other companies listed in PitchBook (see Table B.4). While the main results persist in most specifications, their significance declines, likely due to the smaller sample.

Some deals lack next-round valuation data in PitchBook. Our baseline analysis treats those deals' MPKs as missing. For robustness, we repeat the analysis in Table 5 after imputing zeros for missing next-round valuations, recognizing that many of these startups have failed. See Table B.5 for results and additional details. The results are slightly weaker but remain statistically significant in all but one specification. Some weakening is expected, as imputing a zero for a missing valuation injects measurement error in some cases.

We explore robustness to 14 variations on the data-cleaning steps we take to compute our diligence proxy. Those variations include changing the look-back period for meetings from 18 months to 12 or 24 months before the deal; in meetings with multiple VC devices present, computing the sum or average duration instead of the maximum duration; changing the minimum required meeting length from 10 minutes to either 30 or 60 minutes; and considering meetings only at the VC office or only at the startup office. Table B.6 shows that the baseline coefficient from Table 5 remains statistically significant in all 14 variations.

Table B.7 shows six variations on our extensive-margin analysis from Section 7. We continue to find that startup quality increases monotonically across our five groups if we drop or use different fixed effects; change the cutoffs used to define top VC and long due diligence; or measure startup quality either by the amount raised in the current round (our model-implied proxy for startup quality) or the amount raised in the next round.

9. Conclusion

This paper provides new empirical insights into how VCs choose the intensity of their due diligence. By using smartphone signal data to measure the duration of pre-investment meetings, we show that the due-diligence choice responds not just to startup characteristics but also to VC characteristics and market conditions. Specifically, we show that less due diligence is associated with geographic distance, investor busyness, and hotter deals, sectors, and markets.

We also study the implications of due diligence for capital allocation and investment outcomes. We find that less due diligence is associated with more dispersed investment outcomes, consistent with a model in which VCs optimize learning subject to costs. Viewed through this model, our evidence suggests that due diligence improves the allocation of venture capital to startups. A simple calibration shows that these improvements can be economically large. Patterns in due diligence help predict startup quality. Together, our results indicate that pre-investment effort is an important channel for value creation in VC.

Due diligence is under-researched relative to its importance in practice, and our study takes just one step forward. Our measure captures only a portion of VC due diligence, and the proxy we use is noisy. More work on measuring due diligence is clearly needed. More research on due diligence is also needed outside VC, especially in M&A, private equity, real estate, and other asset classes where diligence plays a major role. Also, this paper does not analyze causality. Identifying exogenous factors that influence due diligence and examining their effects on investment outcomes is another important direction for future work.

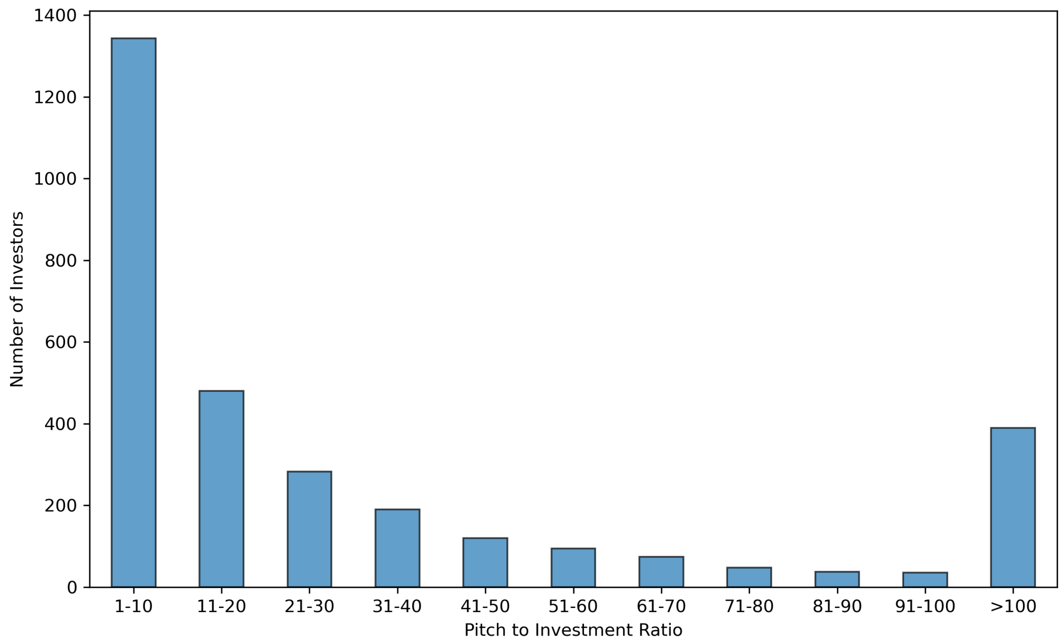

Figure 1. Number of pitches VCs receive per investment. This figure shows the distribution of the pitch-to-investment ratio for investors. The ratio is calculated as the number of captured pitches between a focal investor and any startup from 2018 to 2023, divided by the total number of investments made by that investor during the same period. Only investors with at least one lead investment during the sample period are included.

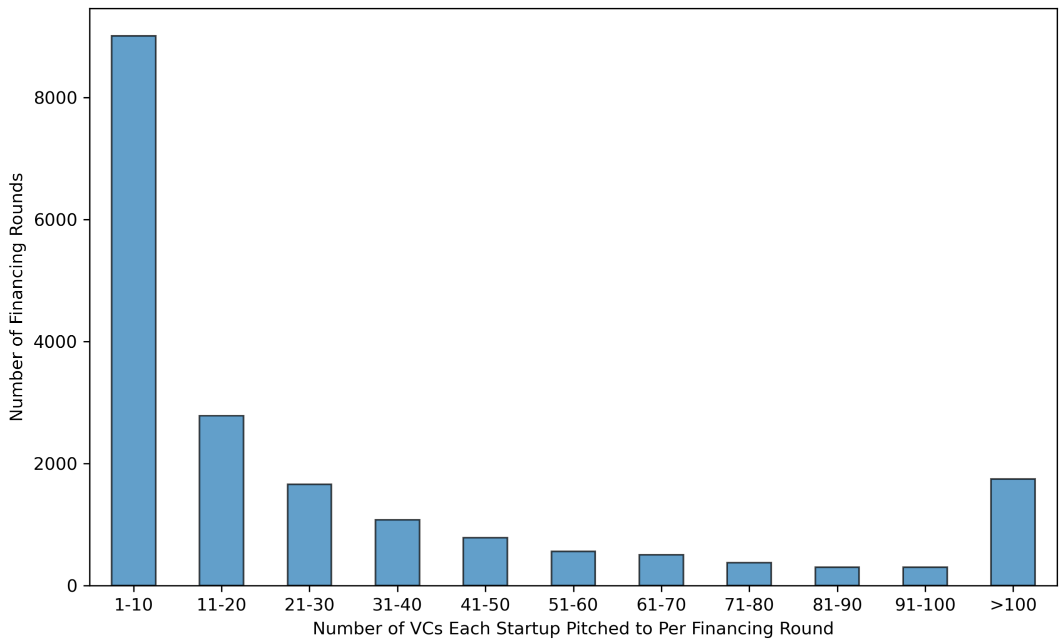

Figure 2. Number of pitches startups make per funding round. This figure shows the distribution of the number of investors that a startup pitches to per funding round during the sample period from 2018 to 2023. Each funding round is treated as an individual observation. The average number is 27, with the 25th, 50th (median), and 75th percentiles at 2, 8, and 30, respectively. To provide a cleaner comparison to the First Round Capital survey, we only include investors classified as VCs, accelerators, or incubators. To mitigate potential data truncation, we exclude the first financing round (and any associated pitch sessions) if a startup's first financing round in the sample period occurred before July 2019, ensuring an 18-month look-back window. For each subsequent financing round, we match pitches that occurred either within 18 months before the deal date or between the previous and current financing rounds, whichever period is shorter.

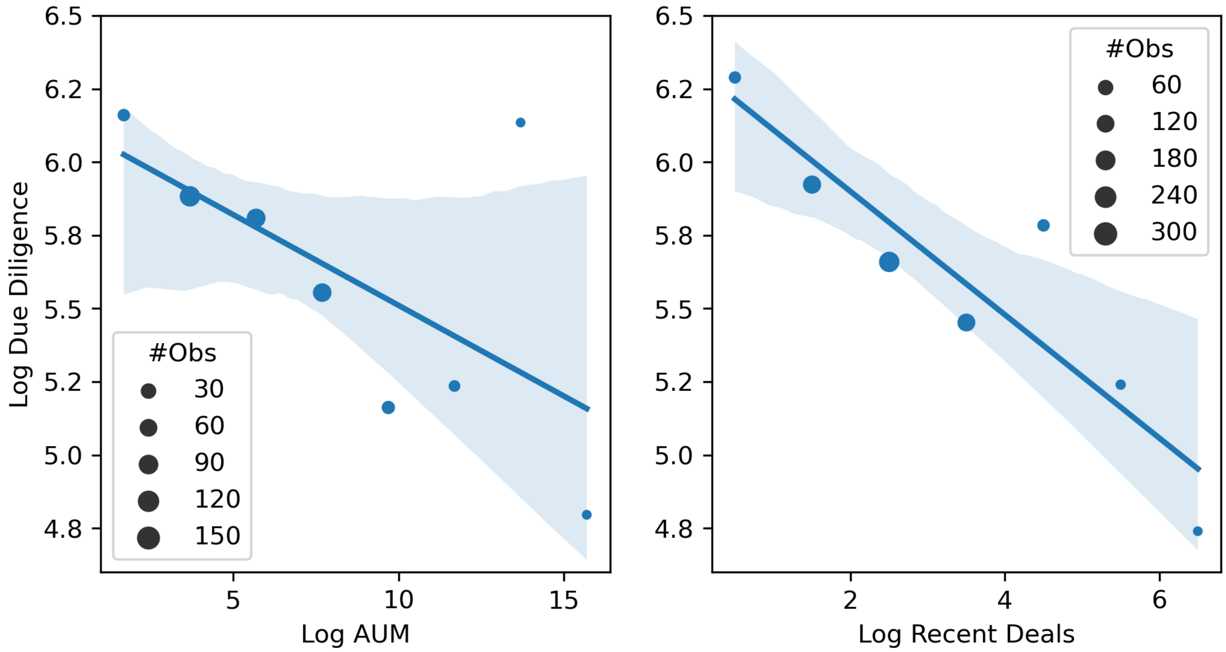

Figure 3. Due diligence and investor characteristics. This figure shows the relation between due diligence duration and investor characteristics. Both subplots use bin-scatter plots, where dot size represents the number of investor-startup observations. The left subplot shows the log of the PitchBook variable 'AUM', measured in millions of dollars. The right subplot shows the log of the number of recent deals, defined as the number of deals the investor invested in within 18 months of the focal deal. The y-axis is the log of due diligence duration, measured in minutes, where 4 stands for approximately 1 hour, 5 stands for approximately 2 hours, and 6 stands for approximately 7 hours. The sample includes only new lead investors with positive due diligence duration. The bands around the fitted line show the 95% confidence intervals, estimated using bootstrap.

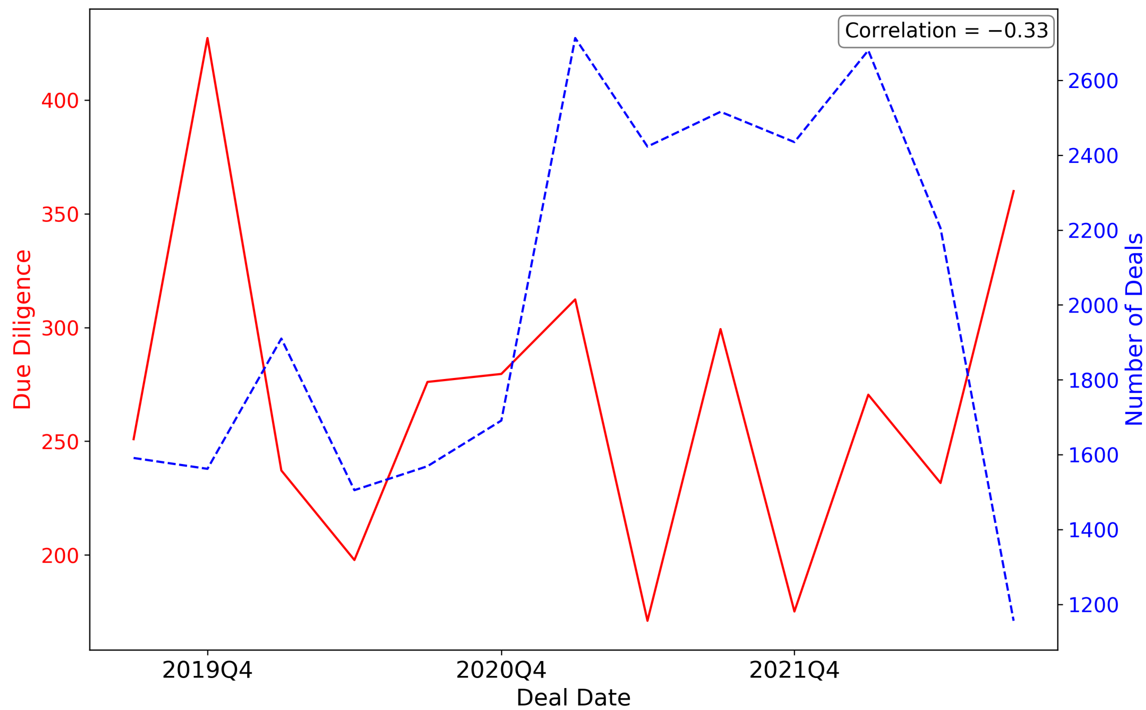

Figure 4. Due diligence and deal volume. This figure shows the due diligence duration and deal volume over time. The solid red line (left y-axis) shows the median due diligence duration, conditional on a positive value, for deals invested in each quarter. Due diligence duration is measured as the total number of minutes that an investor and startup spent together at either the investor's or startup's buildings within 18 months before the investment date. To avoid data truncation issues, the first 18 months are excluded. The dashed blue line (right y-axis) shows the total number of deals categorized as 'Early Stage VC' or 'Later Stage VC' invested in each quarter.

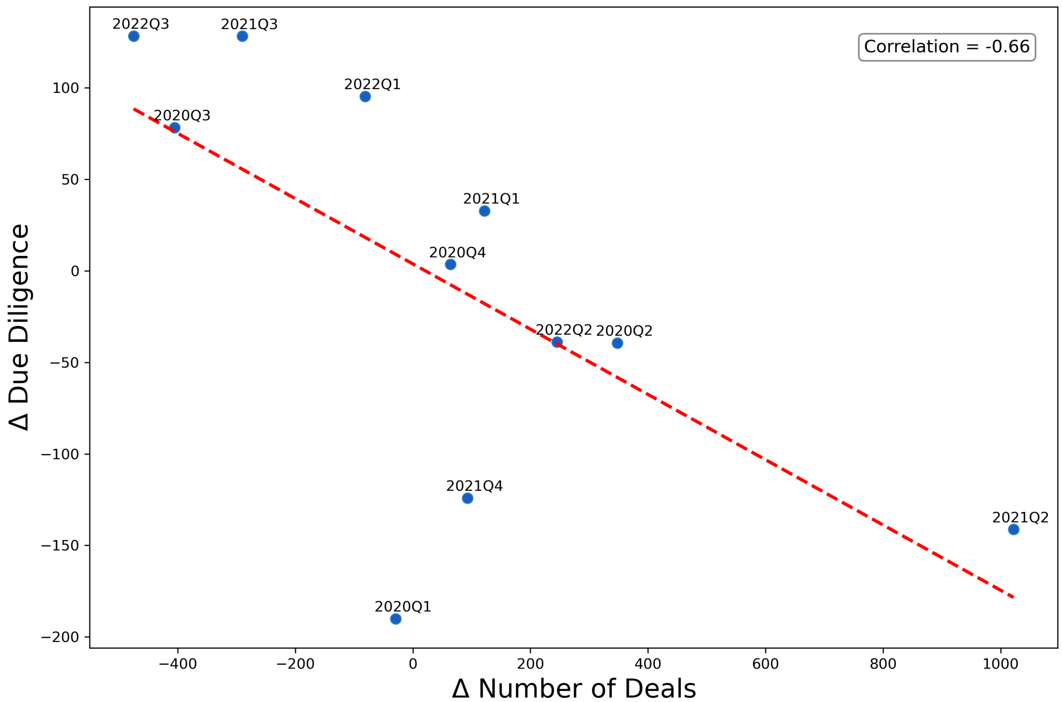

Figure 5. Quarterly changes in due diligence and deal volume. This figure illustrates the relation between the first differences in due diligence duration and deal volume across quarters. The x-axis represents the change in the number of VC deals from the previous quarter to the current quarter, shifted forward by one quarter to better align with market conditions during the due diligence period. The y-axis shows the change in median due diligence minutes over the same two quarters. Each point is labeled with the corresponding quarter, and the fitted trend line is displayed as a red dashed line.

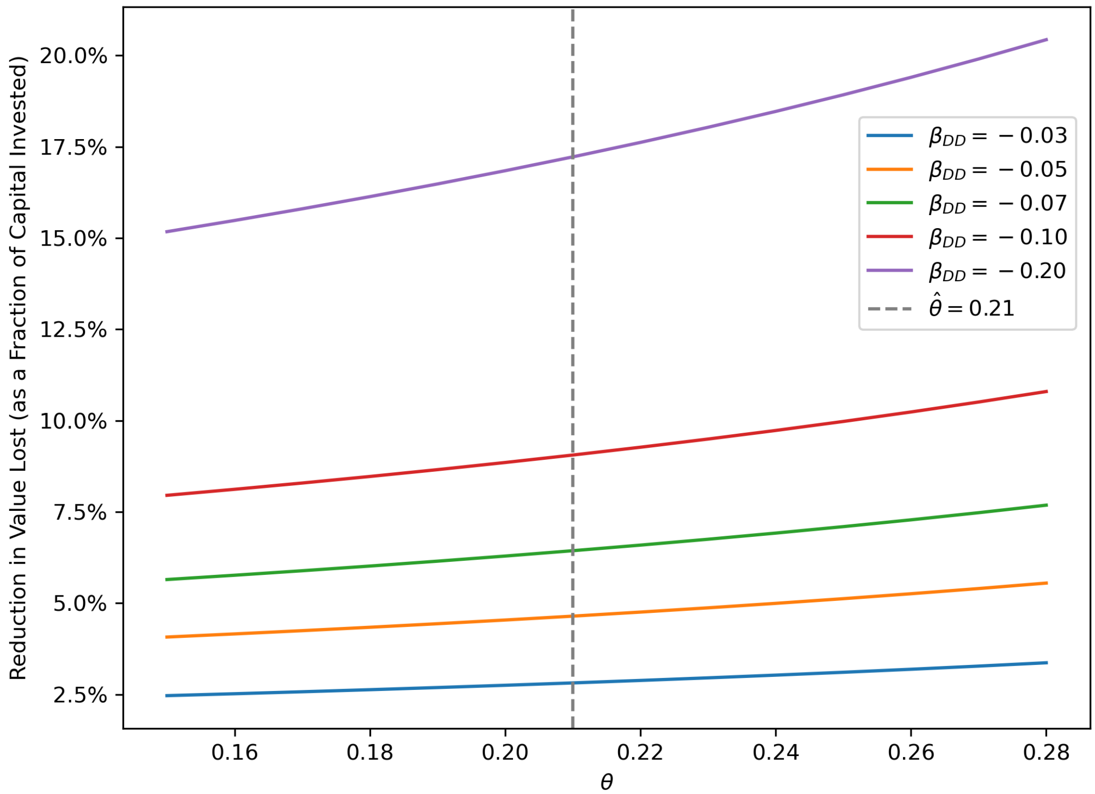

Figure 6. Due diligence and improvements to capital allocation. This figure illustrates how much the value lost due to imperfect information can be reduced by doubling due diligence hours, based on various regression coefficients from Table 5 and different values of the curvature parameter θ. The vertical dashed line denotes the estimate of θ, 0.21. The x-axis ranges from 0.15 to 0.28, which corresponds to the 95% confidence interval for $\hat{\theta}$. The five curves correspond to different β_{DD} coefficient estimates from Panels A, B, and C of Table 5. The y-axis shows the potential reduction in value lost by doubling due diligence, computed using equation (18) for two values of σ^2.

Table 1
Summary statistics

This table presents the summary statistics for the main variables in this study. Panel A presents full sample data. The construction of the full sample is detailed in Section 2.3. Panel B includes only deals with positive due diligence duration. Due diligence is measured as the time that an investor and a startup spend together at either the investor's or startup's buildings within 18 months before the investment date. Marginal Product of Capital (MPK) is computed as the adjusted valuation step-up divided by the current round's invested capital. VC Contacts per Month is the number of other VCs meeting with the focal startup per month within a specific time window, defined as the shorter period between the focal VC's first pitch date and the investment date, or 18 months. When calculating the number of other VCs, we exclude the focal VC and all existing investors from previous rounds. Abnormal Deal Volume is the deal volume for the same year, stage, and industry as the focal deal, divided by the average deal volume over the previous two years. Distance is the distance between the investor's office and the startup's office, measured in kilometers. Deals per Partner is defined as the number of deals the investor invests in within 18 months of the focal deal's investment date, divided by the number of investment professionals at the investor firm, including positions such as principals, partners, directors, and associates, but excluding roles like accountants and marketing directors. Capital Invested is the amount of capital newly injected into the startup in the current round. Lastly, Startup Age is the duration from the startup's founding year to the deal date.

Variables	#Deals	P25	Median	P75	Mean	Std
Panel A: Full Sample						
Due Diligence (hours)	21,655	0.00	0.00	0.00	1.50	18.40
Marginal Product of Capital	5,514	1.66	4.34	10.18	22.79	207.82
VC Contacts per Month	9,977	0.32	1.00	2.89	2.29	3.20
Abnormal Deal Volume	11,410	1.10	1.24	1.61	1.36	0.39
Distance (kilometers)	21,655	41.89	706.72	2826.88	1475.12	1597.38
Deals per Partner	21,160	0.80	2.20	6.00	9.53	22.94
Capital Invested ($millions)	16,325	0.38	4.15	16.50	20.19	62.63
Startup Age (years)	21,220	1.00	3.00	5.00	3.94	4.41
Panel B: Positive DD Sample						
Due Diligence (hours)	1,015	1.28	4.33	19.35	32.04	79.06
Marginal Product of Capital	233	1.42	4.23	9.27	12.80	42.90
VC Contacts per Month	774	0.90	2.68	6.46	4.48	5.15
Abnormal Deal Volume	526	1.05	1.24	1.58	1.35	0.39
Distance (kilometers)	1,015	0.82	2.54	6.40	37.84	268.18
Deals per Partner	982	0.67	2.00	5.00	5.16	13.26
Capital Invested ($millions)	746	0.28	3.00	10.00	12.99	29.16
Startup Age (years)	986	1.00	3.00	5.00	4.02	5.14

Table 2
Comparing due diligence across subsamples

This table summarizes our due diligence measure in different subsamples. ‘Conditional Median’ and ‘Conditional Mean’ reflect statistics for deals with positive due diligence hours, while ‘Unconditional Mean’ includes all deals. ‘Pct > 0’ shows the percentage of deals with positive due diligence hours. For the last four columns, we test the statistical significance of differences between the first subgroup and each of the other subgroups within each panel (e.g., in Panel A, we test the difference between ‘Early Stage VC’ and each of the other three subgroups). Statistical significance is indicated by *, **, and *** at the 10%, 5%, and 1% levels, respectively. We use T-tests to compare group means in the ‘Pct > 0,’ ‘Conditional Mean,’ and ‘Unconditional Mean’ columns. For the ‘Conditional Median’ column, we use Mann–Whitney U tests to compare medians. In Panel A, deal type is based on the PitchBook variable ‘Deal Type.’ In Panel B, industry is based on the PitchBook variable ‘Primary Industry Sector.’ In Panel C, investor location is based on the PitchBook variable ‘HQ Location.’ In Panel D, investor type is based on the PitchBook variable ‘Primary Investor Type,’ with hedge funds grouped under asset managers, and ‘Others’ including categories like ‘Not-For-Profit VC’ and ‘University,’ among 14 additional types.

Subsample	#Deals	Median Deal Size ($Million)	Due Diligence (Hours) Pct >0	Conditional Median	Conditional Mean	Unconditional Mean
Panel A: Deal Type						
Early Stage VC	5100	12.5	4.37	4.32	34.63	1.51
Later Stage VC	4657	22.0	4.23	4.40	32.34	1.37
Accelerator/Incubator	7040	0.1	4.29	4.25	29.10	1.25
Seed Round	4858	3.0	6.03***	4.44	32.89	1.98
Panel B: Startup Industry						
IT	9632	4.6	4.70	4.12	31.23	1.47
Healthcare	4627	6.0	4.28	4.83	42.12	1.80
B2C	3269	2.3	5.90***	4.18	26.16	1.54
B2B	2861	3.0	4.16	4.14	28.77	1.20
Financial Services	571	7.0	4.73	2.76	27.68	1.31
Energy	381	3.0	3.67	4.87	24.79	0.91
Materials and Resources	314	1.5	3.50	9.80*	42.27	1.48
Panel C: Investor Location						
California	8398	5.2	2.08	2.88	16.74	0.35
New York	2697	11.0	8.05***	3.54**	27.49	2.21***
Boston	1214	10.0	5.35***	4.42**	37.40*	2.00***
Others	9346	2.0	5.97***	5.96***	37.98***	2.27***
Panel D: Investor Type						
VC	11360	5.4	4.90	4.45	34.64	1.70
Accelerator/Incubator	6314	0.1	4.61	4.78	29.99	1.38
Growth/Expansion	1136	28.0	3.61**	3.03	13.24***	0.48***
CVC	689	12.0	3.19**	2.38	24.80	0.79**
PE/Buyout	681	32.0	4.55	5.75	42.50	1.93
Government	448	0.3	0.89***	14.81	31.63	0.28***
Asset Manager	268	55.7	3.73	3.19	13.55*	0.51***
Angel Group	241	1.6	7.88*	4.40	22.36	1.76
Family Office	124	4.8	10.48**	4.94	11.57***	1.21
Impact Investing	112	11.4	3.57	2.97	120.92	4.32
Others	282	5.0	8.16**	4.50	33.58	2.74

Table 3
Correlation between due diligence and deal volume

This table presents time-series correlations between aggregate measures of due diligence and VC deal volume. To avoid data truncation issues, the first 18 months of the sample period are excluded. VC deal volume is measured as the total number of 'Early Stage VC' or 'Later Stage VC' deals invested in each quarter. To mitigate the impact of outliers, we use the median of our due diligence measure, conditional on positive values, computed across deals invested in that quarter. However, due diligence often occurs several months before the investment, creating a potential timing mismatch between VC deal volume and due diligence duration. To correct this mismatch, we also include a version where deal volume is shifted forward by one quarter to better align with the market conditions during the due diligence period. We report correlations for both the levels and first differences of the original and shifted samples. Panel A includes all available quarters, while Panel B excludes quarters heavily impacted by COVID-19, which likely disrupted in-person meetings. Specifically, we exclude all quarters before 2021q1, when the first vaccines became widely available. Finally, we estimate the p-value for the correlation by using the p-value of the regression coefficients with a Newey-West adjustment. We follow the standard practice of setting the number of Newey-West lags to the smallest integer greater than or equal to $T^{1/4}$, where T is the number of observations (Green, 2003, p. 267). The resulting number of lags is two quarters. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Level	First-Diff
Panel A. All Quarters		
Original	-0.33	-0.06
Shifted	-0.45**	-0.66***
Panel B. After 2021Q1		
Original	-0.62**	-0.06
Shifted	-0.85***	-0.84***

Table 4
Due diligence, competition, and direct costs

This table shows the relation between due diligence and proxies for due diligence costs. The dependent variable is the log of our due diligence measure. All panels exclude observations with zero diligence duration. Distance denotes the geographic distance between the investor's and startup's office buildings. VC Contacts per Month is the number of other VCs meeting with the focal startup per month within a specific time window, which is the shorter period between the focal VC's first pitch date and the investment date, or 18 months. When calculating the number of other VCs, we exclude the focal VC and all existing investors from previous rounds. Abnormal Deal Volume is the deal volume for the same year, stage, and industry as the focal deal, divided by the average deal volume over the previous two years. Deals per Partner is defined as the number of deals the investor invested in within 18 months of the focal deal's investment date, divided by the number of investment professionals at the investor firm, including positions such as principals, partners, directors, and associates, but excluding roles like accountants and marketing directors. Stage is based on the PitchBook variable 'Deal Type 2,' which takes on values 'Seed Round,' 'Series A,' 'Series B,' and so on; we combine Series D and later rounds into a single category. Industry is based on the PitchBook variable 'Primary Industry Sector,' which includes the seven categories listed in Table 2, Panel B. In all regressions, standard errors, clustered by industry, are shown in parentheses. *, ** and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Baseline

	(1)	(2)	(3)	(4)	(5)
log(VC Contacts per Month)	-0.225**	-0.226**	-0.213**	-0.193**	-0.200**
	(0.083)	(0.071)	(0.077)	(0.067)	(0.077)
log(Distance)	-0.632***	-0.649***	-0.647***	-0.652***	-0.689***
	(0.021)	(0.031)	(0.033)	(0.032)	(0.019)
Observations	774	773	773	773	691
Adjusted R^2	0.226	0.229	0.231	0.230	0.260
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Panel B: Alternative proxy for competition

	(1)	(2)	(3)	(4)	(5)
Abnormal Deal Volume	-0.412**	-0.694*	-0.995**	-1.016**	-1.039*
	(0.122)	(0.340)	(0.354)	(0.369)	(0.430)
log(Distance)	-0.618***	-0.624***	-0.623***	-0.633***	-0.712***
	(0.052)	(0.070)	(0.069)	(0.066)	(0.055)
Observations	526	525	525	525	432
Adjusted R^2	0.202	0.207	0.215	0.214	0.217
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Panel C: Alternative proxy for direct costs

	(1)	(2)	(3)	(4)	(5)
log(VC Contacts per Month)	-0.146*	-0.137*	-0.124	-0.058	-0.053
	(0.072)	(0.070)	(0.073)	(0.087)	(0.099)
log(Deals per Partner)	-0.289**	-0.329**	-0.324**	-0.405***	-0.351***
	(0.113)	(0.106)	(0.102)	(0.100)	(0.072)
Observations	754	753	753	753	671
Adjusted R^2	0.022	0.023	0.024	0.027	0.028
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Table 5
Due diligence and the dispersion in MPK

This table shows results from deal-level regressions with dependent variable equal to $Var(\log(MPK))$, the variance of log MPK. MPK for each deal is computed following equation (12). The fraction's numerator is the financing round's valuation step-up, according to equation (1). Therefore, we measure MPK as the round's valuation step-up divided by the amount invested in the round. This measure equals MPK up to the proportional constant θ, which we assume is equal across observations. We measure the valuation step-up as $PreMV_{next}/(1+r) - PreMV_{current}$. $PreMV_{next}$ refers to the pre-money valuation in the startup's subsequent financing round, and $PreMV_{current}$ is the pre-money valuation in the current round. r is the stock market index return for the corresponding industry and time period, and $K_{current}$ is the current round's invested capital. Throughout this paper, log transformation means $f(x) = \log(1 + x)$ to avoid taking the logarithm of zero. In Panel A, we report the baseline specification, where $Var(\log(MPK))$ is computed as follows. First, we run a regression of deal i's $\log(MPK_i)$ against $\log(DD_i)$ and store the fitted value as $\overline{\log(MPK_i)}$. Then, $Var(\log(MPK_i))$ is computed as $(\log(MPK_i) - \overline{\log(MPK_i)})^2$. In Panel B, we adjust the baseline model by adding control variables: log(Startup Age), 'Pct of Deals in Same Industry', log(K), and 'Time to Next Round'. 'Startup Age' is the time from the startup's founding year to the deal date. 'Pct of Deals in Same Industry' is the proportion of a VC's recent deals (within 18 months) that share the same industry sector as the focal deal. K is the amount of capital newly injected into the startup. 'Time to Next Round' is the duration between the current round and the subsequent round, serving as a proxy for ex-post shocks' variance. These controls are included both in the first step to generate the fitted value and in the second step in Panel B. In Panel C, we adjust the baseline by only using deals with positive due diligence duration. In all regressions, standard errors clustered by industry are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Full sample

	(1)	(2)	(3)	(4)	(5)
log(DD)	-0.066***	-0.047**	-0.047**	-0.034**	-0.019
	(0.018)	(0.014)	(0.014)	(0.013)	(0.018)
Observations	5206	5201	5201	5201	5148
Adjusted R^2	0.000	0.019	0.019	0.169	0.164
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Panel B: Full sample with controls

	(1)	(2)	(3)	(4)	(5)
log(DD)	-0.058***	-0.052***	-0.053***	-0.037***	-0.027*
	(0.009)	(0.008)	(0.008)	(0.008)	(0.012)
log(Startup Age)	0.129**	0.137**	0.129**	-0.065	-0.096*
	(0.038)	(0.045)	(0.045)	(0.048)	(0.043)
Pct of Deals in Same Industry	-0.266***	-0.287***	-0.494***	-0.375***	-0.360***
	(0.063)	(0.063)	(0.110)	(0.068)	(0.076)
log(K)	-0.426***	-0.430***	-0.419***	-0.421***	-0.425***
	(0.028)	(0.029)	(0.031)	(0.020)	(0.022)
Time to Next Round	0.024***	0.025***	0.025***	0.030***	0.031***
	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)
Observations	5044	5040	5040	5040	4990
Adjusted R^2	0.047	0.055	0.056	0.103	0.104
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Stage FE	No	No	No	No	Yes

Panel C: Positive DD sample

	(1)	(2)	(3)	(4)	(5)
log(DD)	-0.103	-0.178***	-0.175***	-0.184*	-0.227***
	(0.060)	(0.047)	(0.046)	(0.080)	(0.049)
Observations	223	216	216	216	158
Adjusted R^2	0.001	-0.008	-0.039	0.070	0.103
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Table 6
Due diligence and the level of MPK

This table shows the relation between due diligence and the level of MPK. The dependent variable is the log of the deal's MPK. Remaining details are the same as in the previous table.

Panel A: Full sample

	(1)	(2)	(3)	(4)	(5)
log(DD)	-0.021	-0.014	-0.014	-0.013	-0.012
	(0.012)	(0.009)	(0.009)	(0.008)	(0.008)
Observations	5206	5201	5201	5201	5148
Adjusted R^2	0.000	0.041	0.054	0.250	0.265
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Panel B: Full sample with controls

	(1)	(2)	(3)	(4)	(5)
log(DD)	-0.035**	-0.030**	-0.029**	-0.017	-0.015
	(0.011)	(0.009)	(0.009)	(0.011)	(0.010)
log(Startup Age)	-0.125**	-0.135**	-0.138**	-0.275***	-0.292***
	(0.043)	(0.045)	(0.047)	(0.039)	(0.038)
Pct of Deals in Same Industry	-0.289**	-0.285**	-0.230	-0.151	-0.159
	(0.105)	(0.115)	(0.144)	(0.163)	(0.175)
log(K)	-0.316***	-0.313***	-0.309***	-0.309***	-0.304***
	(0.025)	(0.022)	(0.023)	(0.016)	(0.014)
Time to Next Round	-0.023**	-0.020**	-0.020**	-0.016*	-0.016*
	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)
Observations	5044	5040	5040	5040	4990
Adjusted R^2	0.165	0.195	0.198	0.313	0.329
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Panel C: Positive DD sample

	(1)	(2)	(3)	(4)	(5)
log(DD)	-0.014	-0.056***	-0.057***	-0.065***	-0.067***
	(0.010)	(0.007)	(0.007)	(0.015)	(0.010)
Observations	223	216	216	216	158
Adjusted R^2	-0.004	0.054	0.052	0.117	0.274
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Table 7
Due diligence and return volatility

This table shows the relation between due diligence and return volatility. The dependent variable is the variance of the log return. The log return is the log of $R_i = \frac{PreMV_{next}}{PreMV_{current}+K_{current}}$, where $PreMV_{next}$ is the next round's pre-money valuation, $PreMV_{current}$ is the current round's pre-money valuation, and $K_{current}$ is the current round's invested capital. In Panel A, we report the baseline specification, where a deal's variance of log return is computed as follows: First, we run a regression of $\log(R_i)$ against $\log(DD_i)$ and store the fitted value as $\overline{\log(R_i)}$. Then, $Var(\log(R_i))$ is computed as $(\log(R_i) - \overline{\log(R_i)})^2$ and used as the dependent variable. Remaining details are the same as in Table 5.

Panel A: Full sample

	(1)	(2)	(3)	(4)	(5)
log(DD)	-0.030***	-0.025***	-0.025***	-0.022**	-0.015*
	(0.006)	(0.003)	(0.003)	(0.006)	(0.007)
Observations	5514	5509	5509	5509	5457
Adjusted R^2	0.000	0.003	0.003	0.038	0.034
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Panel B: Full sample with controls

	(1)	(2)	(3)	(4)	(5)
log(DD)	-0.029***	-0.027***	-0.027***	-0.021**	-0.015
	(0.007)	(0.007)	(0.007)	(0.008)	(0.011)
log(Startup Age)	-0.013	-0.001	0.002	-0.038	-0.064*
	(0.022)	(0.024)	(0.025)	(0.033)	(0.030)
Pct of Deals in Same Industry	-0.053	-0.059	-0.003	0.022	0.025
	(0.098)	(0.097)	(0.100)	(0.086)	(0.085)
log(K)	-0.148***	-0.160***	-0.163***	-0.112***	-0.110***
	(0.022)	(0.018)	(0.015)	(0.024)	(0.021)
Time to Next Round	0.015**	0.017***	0.017***	0.019***	0.019***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	5338	5334	5334	5334	5285
Adjusted R^2	0.018	0.021	0.022	0.038	0.039
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Stage FE	No	No	No	No	Yes

Panel C: Positive DD sample

	(1)	(2)	(3)	(4)	(5)
log(DD)	-0.032	-0.060	-0.056	-0.055	-0.155***
	(0.019)	(0.036)	(0.034)	(0.052)	(0.019)
Observations	233	227	227	227	165
Adjusted R^2	-0.003	-0.000	-0.022	0.001	0.050
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Table 8
Due diligence and the level of returns

This table presents the relation between due diligence duration and investment return. The dependent variable is the deal's log return, defined in the previous table. All remaining details are the same as in Table 5.

Panel A: Full sample

	(1)	(2)	(3)	(4)	(5)
log(DD)	-0.002	-0.003	-0.003	-0.006	-0.009
	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)
Observations	5514	5509	5509	5509	5457
Adjusted R^2	-0.000	0.031	0.039	0.106	0.114
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Panel B: Full sample with controls

	(1)	(2)	(3)	(4)	(5)
log(DD)	-0.009	-0.010	-0.010	-0.008	-0.010
	(0.009)	(0.008)	(0.008)	(0.009)	(0.010)
log(Startup Age)	-0.179***	-0.193***	-0.197***	-0.206***	-0.212***
	(0.019)	(0.023)	(0.025)	(0.023)	(0.024)
Pct of Deals in Same Industry	-0.028	-0.024	-0.099	-0.091	-0.089
	(0.078)	(0.090)	(0.084)	(0.090)	(0.097)
log(K)	-0.142***	-0.136***	-0.131***	-0.104***	-0.105***
	(0.009)	(0.007)	(0.011)	(0.009)	(0.010)
Time to Next Round	-0.003	-0.003	-0.002	-0.002	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	5338	5334	5334	5334	5285
Adjusted R^2	0.084	0.112	0.119	0.129	0.138
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Panel C: Positive DD sample

	(1)	(2)	(3)	(4)	(5)
log(DD)	0.000	-0.038**	-0.042**	-0.046**	-0.040*
	(0.005)	(0.013)	(0.014)	(0.017)	(0.016)
Observations	233	227	227	227	165
Adjusted R^2	-0.004	0.012	0.036	0.143	0.257
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Table 9
Extensive margin analysis

This table shows how due diligence from top VCs predicts startup quality. We estimate a regression with dependent variable equal to an indicator for whether the startup raises a subsequent financing round or exits after the current round. Observations are at the startup-by-round level. The independent variable is a categorical variable for whether the round belongs to one of five groups. Groups are defined based on (1) whether a top VC performs zero, short, or long due diligence (DD) on the deal; and (2) whether a top VC leads the round. Top VCs refer to VC investors in the top 20% by AUM. Long DD is defined as DD above the top 20% of positive DD values, and short DD is defined as DD that is strictly positive and yet below the bottom 20% of positive DD values. We drop observations that do not fall into these five categories (e.g., cases where DD is positive but lies in the middle rather than in the top or bottom 20%). Group 1 is omitted from the regression. To allow for incomplete deals, we compute DD using the following window: the end month is the round's investment date, and the start month is the latest of the following three: (1) January 2018 (the beginning of the sample period); (2) the investment date of the previous round; and (3) 18 months prior to the current round's investment date. The number of observations is 38,329, and the adjusted R-squared is 0.17. We include month, industry, and stage fixed effects, and we cluster standard errors by industry. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Significance levels in the Estimate column test whether other groups differ from the omitted group (Group 1). To test adjacent group differences, we report the p-value for the difference in coefficient versus the previous group.

Group	Definition	Estimate (Stderr.)	p-value for diff. vs. previous group
1	Top VCs do no DD and do not lead round	0.000	
		(.)	
2	Top VCs do short DD but do not lead round	0.020***	0.004
		(0.004)	
3	Top VCs do long DD but do not lead round	0.048***	0.033
		(0.011)	
4	Top VCs do long DD and lead round	0.137***	0.013
		(0.026)	
5	Top VCs do short DD and lead round	0.183***	0.167
		(0.012)	
	Constant	0.318***	
		(0.001)	

REFERENCES

Ai, Hengjie, Kai Li, and Fang Yang, 2020, Financial intermediation and capital reallocation, *Journal of Financial Economics* 138, 663–686.

Alekseeva, Liudmila, Silvia D. Fontana, Caroline Genc, and Hedieh R. Ranjbar. From in-person to online: The new shape of the VC industry, Working paper, KU Leuven.

Aran, Yifat, and Nizan G. Packin, 2024, Due diligence dilemma, *University of Illinois Law Review*.

Asker, John, Allan Collard-Wexler, and Jan De Loecker, 2014, Dynamic inputs and resource (mis) allocation, *Journal of Political Economy* 122, 1013–1063.

Atkin, David, M. Keith Chen, and Anton Popov, 2025, The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley. Working paper, MIT.

Baum, Joel A.C., and Brian S. Silverman, 2004, Picking winners or building them? Alliance, intellectual, and human capital as selection criteria in venture financing and performance of biotechnology startups, *Journal of Business Venturing* 19, 411–436.

Berk, Jonathan B., and Richard C. Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.

Bernstein, Shai, Xavier Giroud, and Richard R. Townsend, 2016, The impact of venture capital monitoring, *Journal of Finance* 71, 1591–1622.

Bernstein, Shai, Arthur Korteweg, and Kevin Laws, 2017, Attracting early-stage investors: Evidence from a randomized field experiment, *Journal of Finance* 72, 509–538.

Bils, Mark, Peter J. Klenow, and Cian Ruane, 2021, Misallocation or mismeasurement? *Journal of Monetary Economics* 124, S39-S56.

Bonelli, Maxime, 2025, Data-driven investors, *Review of Financial Studies*, forthcoming.

Brown, Gregory W., Oleg R. Gredil, and Preetesh Kantak, 2023, Finding fortune: How do institutional investors pick asset managers? *Review of Financial Studies* 36, 3071–3121.

Catherine, Sylvain, Thomas Chaney, Zongbo Huang, David Sraer, and David Thesmar, 2022, Quantifying reduced-form evidence on collateral constraints, *Journal of Finance* 77, 2143–2181.

Charoenwong, Ben, Yosuke Kimura, Alan Kwan, and Eugene Tan, 2024, Capital budgeting, uncertainty, and misallocation, *Journal of Financial Economics* 153, 103779.

Chen, Henry, Paul Gompers, Anna Kovner, and Josh Lerner, 2010, Buy local? The geography of venture capital, *Journal of Urban Economics* 67, 90–102.

Chen, Jiafeng, and Jonathan Roth, 2024, Logs with zeros? Some problems and solutions, *Quarterly Journal of Economics* 139, 891–936.

Chen, M. Keith, Katherine L. Christensen, Elicia John, Emily Owens, and Yilin Zhuo, 2023, Smartphone data reveal neighborhood-level racial disparities in police presence, *Review of Economics and Statistics*, 1–29.

Chen, M. Keith, Kareem Haggag, Devin G. Pope, and Ryne Rohla, 2022, Racial disparities in voting wait times: Evidence from smartphone data, *Review of Economics and Statistics* 104, 1341–1350.

Chernenko, Sergey, Josh Lerner, and Yao Zeng, 2021, Mutual funds as venture capitalists? Evidence from unicorns, *Review of Financial Studies* 34, 2362–2410.

Choi, Jaewon, Xu Tian, Yufeng Wu, and Mahyar Kargar, 2025, Investor demand, firm investment, and capital misallocation, *Journal of Financial Economics* 168, 104039.

Cochrane, John H., 2005, The risk and return of venture capital, *Journal of Financial Economics* 75, 3–52.

Cong, Lin W., Haoyu Gao, Jacopo Ponticelli, and Xiaoguang Yang, 2019, Credit allocation under economic stimulus: Evidence from China, *Review of Financial Studies* 32, 3412–3460.

Cumming, Douglas, and Simona Zambelli, 2017, Due diligence and investee performance, *European Financial Management* 23, 211–253.

Daley, Brendan, Thomas Geelen, and Brett Green, 2024, Due diligence, *Journal of Finance* 79, 2115–2161.

David, Joel M., Hugo A. Hopenhayn, and Venky Venkateswaran, 2016, Information, misallocation, and aggregate productivity, *Quarterly Journal of Economics* 131, 943–1005.

David, Joel M., Lukas Schmid, and David Zeke, 2022, Risk-adjusted capital allocation and misallocation, *Journal of Financial Economics* 145, 684–705.

David, Joel M., and Venky Venkateswaran, 2019, The sources of capital misallocation, *American Economic Review* 109, 2531–2567.

Davis, Dominic-Madori, and Ron Miller, 2022, All we are saying is give due diligence a chance in 2023, TechCrunch.

Ewens, Michael, and Joan Farre-Mensa, 2020, The deregulation of the private equity markets and the decline in IPOs, *Review of Financial Studies* 33, 5463–5509.

Ewens, Michael, Ramana Nanda, and Matthew Rhodes-Kropf, 2018, Cost of experimentation and the evolution of venture capital, *Journal of Financial Economics* 128, 422–442.

Ewens, Michael and Matthew Rhodes-Kropf, 2015, Is a VC partnership greater than the sum of its partners? *Journal of Finance* 70, 1081–1113.

Fu, Jack X., 2024, How does active involvement benefit investors? Evidence from 85 billion cell phone signals, Working paper, University of Hong Kong.

Gerken, William, Steven Irlbeck, Marcus Painter, and Guangli Zhang, 2025, Watching the watchdogs: Tracking SEC inquiries using geolocation data, Working paper, University of Kentucky.

Gompers, Paul, Will Gornall, Steven N. Kaplan, and Ilya Strebulaev, 2020, How do venture capitalists make decisions? *Journal of Financial Economics* 135, 169–190.

Gompers, Paul, Will Gornall, Steven N. Kaplan, and Ilya A. Strebulaev, 2021, Venture capitalists and COVID-19, *Journal of Financial and Quantitative Analysis* 56, 2474–2499.

Gompers, Paul, Steven N. Kaplan, and Vladimir Mukharlyamov, 2016, What do private equity firms say they do? *Journal of Financial Economics* 121, 449–476.

Gompers, Paul, Anna Kovner, Josh Lerner, and David Scharfstein, 2010, Performance persistence in entrepreneurship, *Journal of Financial Economics* 96, 18–32.

Gornall, Will, and Ilya A. Strebulaev, 2020, Squaring venture capital valuations with reality, *Journal of Financial Economics* 135, 120–143.

Green, William H., 2003, *Econometric Analysis, Fifth Edition*, Prentice Hall.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393–408.

Hochberg, Yael V., Alexander Ljungqvist, and Yang Lu, 2007, Whom you know matters: Venture capital networks and investment performance, *Journal of Finance* 62, 251–301.

Hsieh, Chang-Tai, and Peter J. Klenow, 2009, Misallocation and manufacturing TFP in China and India, *Quarterly Journal of Economics* 124, 1403–1448.

Hsu, David H., 2004, What do entrepreneurs pay for venture capital affiliation? *Journal of Finance* 59, 1805—1844.

Jang, Young Soo, and Steven N. Kaplan, 2025, Venture capital start-up selection, Working paper, Penn State University.

Kaplan, Steven N., and Per Strömberg, 2001, Venture capitalists as principals: Contracting, screening, and monitoring, *American Economic Review* 91, 426–430.

Kaplan, Steven N., and Per Strömberg, 2004, Characteristics, contracts, and actions: Evidence from venture capitalist analyses, *Journal of Finance* 59, 2177–2210.

Korteweg, Arthur, and Morten Sørensen, 2010, Risk and return characteristics of venture capital-backed entrepreneurial companies, *Review of Financial Studies* 23, 3738–3772.

Kuchler, Theresa, Yan Li, Lin Peng, Johannes Stroebel, and Dexin Zhou, Social proximity to capital: Implications for investors and firms, *Review of Financial Studies* 35, 2743–2789.

Lyonnet, Victor, and Léa H. Stern, 2022, Venture capital (mis) allocation in the age of AI, Working paper, Ohio State University.

Mathews, Jessica, and Anna Sraders, 2023, How Tiger Global, one of the biggest backers of startups over the past decade, fell to Earth, Fortune.

Metrick, Andrew, and Ayako Yasuda, 2021, *Venture Capital and the Finance of Innovation, Third Edition,* John Wiley & Sons.

Offenberg, David, and Christo Pirinsky, 2015, How do acquirers choose between mergers and tender offers? *Journal of Financial Economics* 116, 331–348.

PitchBook, 2024, Venture capital monitor Q2 2024.

Plapperer, Felix, 2022, Funding round pre-emption 101, Medium.com.

Sørensen, Morten, 2007, How smart is smart money? A two-sided matching model of venture capital, *Journal of Finance* 62, 2725–2762.

TDK Ventures, 2024, Doing diligence well in venture investing: Going back to the future.

Tyebjee, Tyzoon T., and Albert V. Bruno, 1984, A model of venture capitalist investment activity, *Management Science* 30, 1051–1066.

Wangerin, Daniel, 2019, M&A due diligence, post-acquisition performance, and financial reporting for business combinations, *Contemporary Accounting Research* 36, 2344–2378.

Whited, Toni M., and Jake Zhao, 2021, The misallocation of finance, *Journal of Finance* 76, 2359–2407.

Wilson, Fred, 2006, VCs charging fees, `https://avc.com/2006/07/vcs_charging_fe/`.

Appendix

A.1. Variable definitions

Variable Name	Definition and Explanation
Due Diligence	The total hours investors and startups spent together within 18 months before the investment date, either at the VC's building or the startup's building. Specifically, for meetings at startup buildings, we first identify devices likely belonging to VC employees. If a potential VC employee's device is detected within 200 meters of a startup's office and remains there for at least 10 minutes, we consider it a potential meeting. Finally, we apply several additional filters to reduce false positives. We use the same methods to measure startup employees visiting VC buildings.
Marginal Product of Capital (MPK)	Marginal Product of Capital (MPK) is computed as the adjusted valuation step-up divided by the current round's invested capital. Specifically, $MPK = \frac{PreMV_{next}/(1+r) - PreMV_{current}}{K_{current}}$, where $PreMV_{next}$ is the next round's pre-money valuation, r is the stock market index return for the corresponding industry and time period, $PreMV_{current}$ is the current round's pre-money valuation, and $K_{current}$ is the amount invested in the current round.
Sector Stock Portfolio Return (r)	To compute r, we first assign each startup a sector stock portfolio as follows. We obtain S&P Composite 1500 Index returns from Compustat. Since PitchBook categorizes deals into seven industries, we match each PitchBook 'Primary Industry Sector' category to the closest corresponding index in the S&P Composite 1500: 'Business Products and Services (B2B)' with 'SP1500 Industrials .S,' 'Information Technology' with 'SP1500 Information Tech .S,' 'Healthcare' with 'SP1500 Health Care .S,' 'Energy' with 'SP1500 Energy .S,' 'Consumer Products and Services (B2C)' with 'SP1500 Consumer Staples .S,' 'Materials and Resources' with 'SP1500 Materials .S,' and 'Financial Services' with 'SP1500 Financials .S.' Next, we compute the return on the sector portfolio over the time range between the startup's current and subsequent financing rounds.
VC Contacts per Month	The number of other VCs meeting with the focal startup per month within a specific time window, defined as the shorter period between the focal VC's first pitch date and the investment date, or 18 months. When calculating the number of other VCs, we exclude the focal VC, all co-investors in the current round, and all investors from previous rounds.
Abnormal Deal Volume	The number of deals in the same year, stage, and industry as the focal deal, divided by the average number of deals in the previous two years.

Distance	The distance between the investor's office and the startup's office, measured in kilometers.
Deals per Partner	The number of deals the investor invests in within 18 months of the focal deal's investment date, divided by the PitchBook variable 'Investment Professional Count,' which represents the number of investment professionals at the investor firm, including positions such as principals, partners, directors, and associates, but excluding roles like accountants and marketing directors.
Capital Invested	The amount of capital newly injected into the startup in the current round, measured in millions of dollars.
Startup Age	The duration from the startup's founding year to the deal date.
Assets Under Management (AUM)	The PitchBook variable 'AUM' represents the amount of capital managed by an investor.
Exit Rate	The ratio of the PitchBook variable 'Total Exits' to the PitchBook variable 'Total Investments' for the focal investor.
Number of Recent Deals	The number of deals the investor invested in within 18 months of the focal deal's investment date.
Recent Deals per Employee	The number of recent deals divided by the number of investment professionals at the investor firm.
Time to Next Round	The duration between the current round and the subsequent round, serving as a proxy for ex-post shocks.
Pct of Deals in Same Industry	The proportion of a VC's recent deals (within 18 months) that share the same industry sector as the focal deal.
Investment Return	The log return of a deal is calculated by first dividing the pre-money valuation of the next round by the post-money valuation of the current round, and then taking the natural logarithm of the result.
Value Lost	The gap in profits between the current allocation and the optimal allocation, due to imperfect information.
Reduction in Value Lost	The amount of value lost due to imperfect information that can be reduced by doubling due diligence hours.

A.2. Model solution and proofs

Solution: We start by characterizing the Bayesian learning problem as follows:

$$\log a = \mu_0 + \eta, \tag{A.1}$$
$$S = \log a + \delta = \mu_0 + \eta + \delta, \tag{A.2}$$

where $\eta \sim N(0, 1/\nu_0)$ and $\delta \sim N(0, 1/\tau)$ are independent of each other. Note η is observed at $t = 2$ and $\eta + \delta$ is observed at $t = 1$. The posterior mean of $\log a$ can then be expressed as

$$\mu_1 = \mu_0 + \left(\frac{\tau}{\nu_0 + \tau}\right)(\eta + \delta). \tag{A.3}$$

Using properties of the lognormal distribution, we can express the posterior mean of a as

$$\hat{a} = \exp\left(\mu_0 + \left(\frac{\tau}{\nu_0 + \tau}\right)(\eta + \delta) + \frac{1}{2\nu_1}\right). \tag{A.4}$$

Next, we derive the first-order condition for the choice of τ. At $t = 0$, the VC solves

$$\max_{\tau} E_0[W(a, \hat{a}; S)] \tag{A.5}$$
$$W(a, \hat{a}) = a(K^*)^\theta - K^* - c\tau, \tag{A.6}$$

recalling that K^* is a function of $\hat{a}$, which is a function of S. Substituting equation (A.4) into equation (7) yields

$$K^* = \theta^{\frac{1}{1-\theta}} \exp\left[\left(\frac{1}{1-\theta}\right)\left(\mu_0 + \left(\frac{\tau}{\nu_0 + \tau}\right)(\eta + \delta) + \frac{1}{2\nu_1}\right)\right]. \tag{A.7}$$

This term is lognormally distributed given time-zero beliefs, because η and δ are both normally distributed as of $t = 0$. The mean of the argument in square brackets is

$$A \equiv \left(\frac{1}{1-\theta}\right)\left(\mu_0 + \frac{1}{2\nu_1}\right). \tag{A.8}$$

The variance of the argument in square brackets is

$$\begin{aligned} B &\equiv \left(\frac{1}{1-\theta}\right)^2 \left(\frac{\tau}{\nu_0 + \tau}\right)^2 \left(\frac{1}{\nu_0} + \frac{1}{\tau}\right) \\ &= \left(\frac{1}{1-\theta}\right)^2 \frac{\tau}{\nu_0}\left(\frac{\tau}{\nu_0 + \tau}\right). \end{aligned} \tag{A.9}$$

Next, we analyze the first term of equation (A.6):

$$\begin{aligned} a(K^*)^\theta &= \exp\{\mu_0 + \eta\}(K^*)^\theta \\ &= \exp\{\mu_0 + \eta\}\,\theta^{\frac{\theta}{1-\theta}} \exp\left[\left(\frac{\theta}{1-\theta}\right)\left(\mu_0 + \left(\frac{\tau}{\nu_0 + \tau}\right)(\eta + \delta) + \frac{1}{2\nu_1}\right)\right] \\ &= \theta^{\frac{\theta}{1-\theta}} \exp\left\{\mu_0\left(\frac{1}{1-\theta}\right) + \eta\left(1 + \left(\frac{\theta}{1-\theta}\right)\left(\frac{\tau}{\nu_0 + \tau}\right)\right)\right. \\ &\quad \left. + \delta\left(\frac{\theta}{1-\theta}\right)\left(\frac{\tau}{\nu_0 + \tau}\right) + \left(\frac{\theta}{1-\theta}\right)\left(\frac{1}{2\nu_1}\right)\right\}. \end{aligned} \tag{A.10}$$

This term is lognormally distributed given time-zero beliefs. The mean of the term in braces is

$$\begin{aligned} C &\equiv \mu_0 \left(\frac{1}{1-\theta}\right) + \left(\frac{\theta}{1-\theta}\right)\left(\frac{1}{2\nu_1}\right) \\ &= \left(\frac{1}{1-\theta}\right)\left(\mu_0 + \frac{\theta}{2\nu_1}\right). \end{aligned} \tag{A.11}$$

The variance of the term in braces is

$$D \equiv \frac{1}{\nu_0}\left(1 + \left(\frac{\theta}{1-\theta}\right)\left(\frac{\tau}{\nu_0+\tau}\right)\right)^2 + \frac{1}{\tau}\left(\frac{\theta}{1-\theta}\right)^2\left(\frac{\tau}{\nu_0+\tau}\right)^2. \tag{A.12}$$

We then have

$$E_0[W(a, \hat{a}; S)] = \theta^{\frac{\theta}{1-\theta}} \exp\{C + D/2\} - \theta^{\frac{1}{1-\theta}} \exp\{A + B/2\} - c\tau. \tag{A.13}$$

The optimal τ^* can be solved for by taking the first-order condition of expression (A.13) with respect to τ, noting that A,B, C, and D are functions of τ. This first-order condition is

$$\theta^{\frac{\theta}{1-\theta}} \exp\{C + D/2\} \left(\frac{\partial C}{\partial \tau^*} + \frac{1}{2}\frac{\partial D}{\partial \tau^*}\right) - \theta^{\frac{1}{1-\theta}} \exp\{A + B/2\} \left(\frac{\partial A}{\partial \tau^*} + \frac{1}{2}\frac{\partial B}{\partial \tau^*}\right) - c = 0 \tag{A.14}$$

where

$$\frac{\partial C}{\partial \tau^*} + \frac{1}{2}\frac{\partial D}{\partial \tau^*} = \frac{\theta}{2(1-\theta)^2(\nu_0+\tau^*)^2}, \tag{A.15}$$

$$\frac{\partial A}{\partial \tau^*} + \frac{1}{2}\frac{\partial B}{\partial \tau^*} = \frac{\tau^{*2} + \nu_0(\theta - 1 + 2\tau^*)}{2(1-\theta)^2\nu_0(\nu_0+\tau^*)^2}. \tag{A.16}$$

Therefore, we have

$$\begin{aligned} \frac{1}{2(1-\theta)^2\nu_0(\nu_0+\tau^*)^2}\theta^{\frac{1}{1-\theta}} \Big[&\nu_0 \exp\{C + D/2\} \\ &- (\tau^{*2} + \nu_0(\theta - 1 + 2\tau^*) \exp\{A + B/2\} \Big] - c = 0. \end{aligned} \tag{A.17}$$

We can then rewrite the first-order condition for τ in equation (A.17) as

$$f(\tau^*; \theta, \nu_0, \mu_0) - c = 0. \tag{A.18}$$

Proof of Prediction 1: To obtain comparative statics of τ^* with respect to c, we differentiate both sides of the equation (A.18) with respect to c:

$$\begin{aligned} \frac{df}{d\tau^*}\frac{d\tau^*}{dc} - 1 = 0 \\ \frac{df}{d\tau^*}\frac{d\tau^*}{dc} = 1. \end{aligned} \tag{A.19}$$

Therefore, if $\frac{df}{d\tau^*} < 0$, it follows that $\frac{d\tau^*}{dc} < 0$. Differentiating with respect to τ^*, we have

$$\frac{df}{d\tau^*} = -\frac{\theta^{\frac{1}{1-\theta}}}{4(\theta-1)^4\nu_0^2(\tau+\nu_0)^4}\left[E_1 \times \exp\left(C + D/2\right) + (E_2 + E_3) \times \exp\left(A + B/2\right)\right]. \tag{A.20}$$

where

$$E_1 = \nu_0^2\left(4(\theta-1)^2(\tau+\nu_0) - \theta\right) \tag{A.21}$$
$$E_2 = \left(\tau^2 + \nu_0(\theta + 2\tau - 1)\right)^2 \tag{A.22}$$
$$E_3 = 4(\theta-1)^2\nu_0^2(1-\theta+\nu_0)(\tau+\nu_0). \tag{A.23}$$

Note that $E_2 > 0$ and $E_3 > 0$ for $\theta \in (0,1)$, $\tau > 0$, and $\nu_0 > 0$. The sufficient condition for $\frac{df}{d\tau^*} < 0$ is given by

$$\begin{aligned} E_1 = \nu_0^2\left(4(\theta-1)^2(\tau+\nu_0) - \theta\right) &> 0 \\ \tau + \nu_0 &> \frac{\theta}{4(\theta-1)^2} \\ \tau &> \frac{\theta}{4(\theta-1)^2} - \nu_0. \end{aligned} \tag{A.24}$$

Since we have $\tau > 0$, a sufficient condition is

$$\begin{aligned} 0 &\geq \frac{\theta}{4(\theta-1)^2} - \nu_0 \\ \nu_0 &\geq \frac{\theta}{4(\theta-1)^2}. \end{aligned} \tag{A.25}$$

Therefore, if inequality (A.25) is satisfied, it follows that $\frac{df}{d\tau^*} < 0$ and hence $\frac{d\tau^*}{dc} < 0$.

Proof of Prediction 2: From equation (11),

$$\log(MPK) = \log(a) - \log(\hat{a}), \tag{A.26}$$

the variance of which equals $Var(\log(a)|S) = 1/\nu_1 = 1/(\nu_0 + \tau)$, where the last equality uses equation (6).

Proof of Prediction 3: Recall that after seeing S, $\log(a) \sim N(\mu_1, 1/\nu_1)$. By properties of the lognormal distribution,

$$\hat{a} \equiv E[a|S] = \exp(\mu_1 + 1/(2\nu_1)). \tag{A.27}$$

Combining the previous two relations,

$$\log(a) - \log(\hat{a}) \quad \sim \quad N(\mu_1 - (\mu_1 + 1/(2\nu_1)), 1/\nu_1) \tag{A.28}$$
$$\log(a/\hat{a}) \quad \sim \quad N(-1/(2\nu_1)), 1/\nu_1), \tag{A.29}$$

and

$$E[MPK] \quad = \quad E[a/\hat{a}] \tag{A.30}$$
$$\quad = \quad \exp(-1/(2\nu_1) + 1/(2\nu_1)) = 1, \tag{A.31}$$

where the last line again uses properties of the lognormal distribution.

Proofs of claims in Section 6.3:

1. Derivation of equation (17). Expand definition of $\Delta\Pi$:

$$\begin{aligned} \Delta\Pi &= (aK^{*\theta} - K^*) - (aK_i^{**\theta} - K_i^{**}) && \text{(A.32)} \\ &= a(K^{*\theta} - K_i^{**\theta}) - (K^* - K_i^{**}). && \text{(A.33)} \end{aligned}$$

From equations (7) and (15), we can write

$$K_i^{**} = K^* \left(\frac{a}{\hat{a}}\right)^{\frac{1}{1-\theta}}.$$

Substituting that expression into the one above and rearranging,

$$\Delta\Pi = aK^{*\theta}(1 - MPK^{\frac{\theta}{1-\theta}}) - K^*(1 - MPK^{\frac{1}{1-\theta}}). \tag{A.34}$$

Dividing both sides by K^*,

$$\frac{\Delta\Pi}{K^*} = aK^{*(\theta-1)}(1 - MPK^{\frac{\theta}{1-\theta}}) - (1 - MPK^{\frac{1}{1-\theta}}). \tag{A.35}$$

From the FOC for K^*, we know $K^{*(\theta-1)} = (\theta\hat{a})^{-1}$. Substituting into the equation above yields

$$\begin{aligned} \frac{\Delta\Pi}{K^*} &= a(\theta\hat{a})^{-1}(1 - MPK^{\frac{\theta}{1-\theta}}) - (1 - MPK^{\frac{1}{1-\theta}}) && \text{(A.36)} \\ &= \frac{1}{\theta}MPK(1 - MPK^{\frac{\theta}{1-\theta}}) - (1 - MPK^{\frac{1}{1-\theta}}) && \text{(A.37)} \\ &= \frac{1}{\theta}(MPK - MPK^{\frac{1}{1-\theta}}) - (1 - MPK^{\frac{1}{1-\theta}}). && \text{(A.38)} \end{aligned}$$

2. Proof that $\Delta\Pi/K^*$ attains a maximum of zero when $MPK = 1$. Take the derivative of equation (17) w.r.t. MPK:

$$\frac{d}{d\,(MPK)}\frac{\Delta\Pi}{K^*} = \frac{1}{\theta}\left(1 - MPK^{1/(1-\theta)-1}\right). \tag{A.39}$$

That derivative equals zero when $MPK = 1$. From equation (17), $\Delta\Pi/K^* = 0$ when $MPK = 1$. We further show that

$$\frac{d^2}{d\,(MPK)^2}\frac{\Delta\Pi}{K^*} = -\frac{1}{1-\theta}MPK^{1/(1-\theta)-2}, \tag{A.40}$$

which is strictly negative since $\theta \in (0, 1)$ and $MPK > 0$. Therefore, $\Delta\Pi/K^*$ achieves a maximum value of zero at $MPK = 0$.

3. Derivation of equation (18). We start from equation (17). Given that MPK follows a lognormal distribution with $\log(MPK) \sim N(\mu, \sigma^2)$, we aim to find the expectation $E\left(\frac{\Delta\Pi}{K^*}\right)$. The first term involves $E(MPK)$, where

$$E(MPK) = \exp\left(\mu + \frac{\sigma^2}{2}\right).$$

The second term involves $E\left(MPK^{\frac{1}{1-\theta}}\right)$, where $\log(MPK^{\frac{1}{1-\theta}}) \sim N(\frac{\mu}{1-\theta}, \frac{\sigma^2}{(1-\theta)^2})$. Therefore,

$$E\left(MPK^{\frac{1}{1-\theta}}\right) = \exp\left(\frac{\mu}{1-\theta} + \frac{\sigma^2}{2(1-\theta)^2}\right).$$

Substituting these into the original expression yields

$$E\left(\frac{\Delta\Pi}{K^*}\right) = \frac{1}{\theta}\left(\exp\left(\mu + \frac{\sigma^2}{2}\right) - \exp\left(\frac{\mu}{1-\theta} + \frac{\sigma^2}{2(1-\theta)^2}\right)\right) - \left(1 - \exp\left(\frac{\mu}{1-\theta} + \frac{\sigma^2}{2(1-\theta)^2}\right)\right).$$

Combining terms produces equation (18).

A.3. Method for estimating μ, σ^2, and θ

Let $\tilde{V}$ denote the financing round's valuation step-up measured in the data, and $V = aK^\theta$ denote the valuation step-up in the model. The two differ because $\tilde{V}$ includes the effects of ex-post shocks and measurement error that are outside our model. We model those shocks $\tilde{\epsilon}$ as follows:

$$\log(\tilde{V}) = \log(V) + \log(\tilde{\epsilon}),$$

where $E[\log(\tilde{\epsilon})|V] = 0$. Substituting in $V = aK^\theta$, we obtain

$$\log(\tilde{V}) = \theta\log(K) + \log(a) + \log(\tilde{\epsilon}). \tag{A.41}$$

This equation resembles a regression model of $\log(\tilde{V})$ on $\log(K)$, where the error term is $\log(a) + \log(\tilde{\epsilon})$. Typically, estimating this regression using OLS results in a biased estimator, as the investment choice is endogenous to the error term. Mitigating that concern in our setting, the shock $\tilde{\epsilon}$ hits the firm after K is chosen, so $E[\log(\tilde{\epsilon})|K] = 0$ is plausible. We partially address the endogeneity of K w.r.t. $\log a$ by first noting that $\log a$ is not directly observed when K is chosen, and second by soaking up as much variation in $\log a$ by including startup and time fixed effects in our regression. We proceed with the following deal-level OLS model:

$$\log(\tilde{V}) = \beta_0 + \beta_1\log(K) + \alpha + \alpha_t + \delta, \tag{A.42}$$

where α and α_t represent startup and month fixed effects, respectively. To input data for $\log(\tilde{V})$, we use panel data from 2018 to 2023 and set the next round valuation to zero for deals made before the year $[2023 - \tau]$ that lack subsequent round information. We set τ equal to three years, given that over 90% of deals in our sample raised the next round within three years. Additionally, we set all negative valuation step-ups to zero to ensure that the calculation remains valid after applying the $\log(1 + x)$ transformation.

Guided by equation (A.41), we estimate θ as the coefficient β_1 in the OLS regression (A.42). We estimate θ to be 0.21, with a 95% confidence interval ranging from 0.15 to 0.28. We show results for all θ values in this confidence interval, as we recognize we have not fully resolved the endogeneity of K w.r.t. $\log(a)$ when estimating (A.42) by OLS.

The next step is to estimate μ, the mean of $\log(MPK)$. By equation (12),

$$MPK = \theta \frac{V}{K} = \theta \frac{\tilde{V}}{K} \cdot \frac{1}{\tilde{\epsilon}}, \tag{A.43}$$

so

$$\log(MPK) = \log(\theta) + \log\left(\tilde{V}/K\right) - \log(\tilde{\epsilon}). \tag{A.44}$$

Thus, the mean of $\log(MPK)$ is given by $\log(\theta) + E\left[\log\left(\tilde{V}/K\right)\right]$, since $E[\log(\tilde{\epsilon})] = 0$. We substitute in the estimate of $\theta = 0.21$ and compute $E\left[\log\left(\tilde{V}/K\right)\right]$ as the sample average log ratio of value step-up to capital invested. The resulting estimate of μ is -0.06.

Next, we focus on estimating the variance of $\log(MPK)$. Rearranging equation (A.44) and taking variances of both sides yields

$$Var(\log(MPK)) = Var(\log(\tilde{V}/K)) - Var(\log(\tilde{\epsilon})). \tag{A.45}$$

We use the variance of the estimated residuals δ from regression (A.42) as an estimator for $Var(\log(\tilde{\epsilon}))$, and we compute $Var(\log(\tilde{V}/K))$ as the sample variance of the ratio of valuation step-up to capital invested. Plugging those values into equation (A.45) produces an estimate of $Var(\log(MPK)) = \sigma^2$ equal to 0.54.

Online Appendix

Contents:

B.1. Model extension with bargaining over the surplus

We extend the model in Section 5 to allow the VC and startup to bargain over the deal's surplus. In reality, this corresponds to bargaining over the size of the VC's ownership stake or, equivalently, over the pre-money valuation. We modify the assumptions as follows. Let V_0 be the fundamental value of the startup's assets in place at $t = 0$, before this financing round. Let aK^θ denote the increase in the startup's fundamental value resulting from this financing round. The VC captures a fraction $\beta > 0$ of the deal's surplus, so β reflects the VC's bargaining power. We replace the objective function in equation (1) with the assumption that the VC maximizes its expected dollar profits from the deal. We also assume a zero discount rate for simplicity.

Given the assumptions above, the startup's fundamental value immediately before the next financing round, at $t = 2$, is $V_0 + aK^\theta$. The startup's value immediately after this financing round, at $t = 1$, is then $V_0 + \hat{a}K^\theta$. The surplus generated by this financing round is $T = \hat{a}K^\theta - K$. Since the VC captures a fraction β of the surplus, the VC's expected dollar profit, gross of due-diligence costs, equals βT. To see this, note that the dollar value of the VC's ownership stake immediately after this round is $K + \beta T$. The VC's expected gross dollar profit equals that value minus the VC's investment, K, which simplifies to βT. Let $\tilde{c} = \beta c$ denote the VC's adjusted due-diligence cost parameter in this extended model. The VC maximizes its expected profit, net of diligence costs:

$$\max_{\tau, K} E[\beta T - \tilde{c}\tau], \tag{B.1}$$

which can be rewritten as

$$\max_{\tau, K} \beta E[aK^\theta - K - c\tau]. \tag{B.2}$$

This objective function is the same as the baseline model's equation (1) except for the positive multiplicative constant β. Therefore, the baseline model and this extension produce the same predicted values of K and τ, for any $\beta > 0$, with one difference: the diligence cost $\tilde{c}$ in the extended model is a factor β below its counterpart c from the baseline model. That difference reflects that the VC in the extended model pays the full diligence cost but captures only a fraction of the surplus.

The VC obtains a fraction ownership stake in the startup equal to

$$\frac{K + \beta T}{V_0 + \hat{a}K^\theta}, \tag{B.3}$$

which equals the value of the VC's stake divided by the startup's value after the deal. From equation (7), $T > 0$. Therefore, the expression above implies that, all else equal, higher bargaining power β implies the VC obtains a larger fraction ownership stake in the startup.

A commonly used valuation metric in VC is the post-money valuation, defined as the dollar amount invested divided by the VC's fraction ownership stake. In our setting,

$$\text{Post-money valuation} = K / \left[\frac{K + \beta T}{V_0 + \hat{a}K^\theta}\right] = (V_0 + \hat{a}K^\theta)\frac{K}{K + \beta T}. \tag{B.4}$$

This expression shows that, all else equal, higher bargaining power β produces a lower post-money valuation. The same result applies to the pre-money valuation, which is simply the post-money valuation minus K. In the equation above, the term $K/(K + \beta T)$ is the wedge between the company's fundamental value after the deal $(V_0 + \hat{a}K^{\theta})$ and the post-money valuation, which is just an accounting construct. This wedge exists because the definition of post-money valuation implicitly assumes the VC captures none of the surplus (i.e., $\beta = 0$).

B.2. Selection bias

Section 6.1 notes that pre-money valuations are more likely to be reported by more-successful startups, raising potential concerns about selection bias. Here, we analyze the selection bias that can occur in the two main relations we study: the relation between the levels of due diligence and MPK, and the relation between the variance of MPK and the level of due diligence.

We simulate the model from Section 5 to illustrate the bias. We simulate values of MPK across a range of diligence levels, corresponding to τ in our model. We assume that startups with the smallest valuation step-ups between rounds—those that are likely to be least successful and have low MPK values—do not report the next round pre-money valuations required to compute MPK. Therefore, low values of MPK are truncated. Simulation results are in the figure below.

Panel A describes the relation between the levels of MPK and due diligence. The low, truncated values of MPK are shown as red dots in the top panel. The blue line shows the best-fit line if we had perfect, non-truncated data. The red line shows the best-fit line if the red data points are missing from our sample. The blue line is approximately flat, consistent with our Prediction 3. The red line slopes down, indicating that truncation biases downward the coefficient of log(MPK) on the level of due diligence. The downward bias occurs because the variance of MPK declines with the level of diligence (Prediction 2 in our model). Therefore, the low, truncated MPK realizations are more common at low levels of diligence. Truncation therefore shifts the best-fit line up at low levels of diligence, inducing a spuriously negative slope in the red line.

Panel B describes the relation between the level of due diligence and the variance of MPK. It plots each observation's value of $Var(\log(MPK_i))$, the squared deviation of MPK from its fitted value. Again, the blue line shows the best-fit line if we had all data, and the red line shows its counterpart if we had only data that suffers from truncation. The red line slopes down less than the blue line, indicating that truncation biases upward (toward zero) the coefficient of $Var(\log(MPK_i))$ on the level of due diligence. Therefore, selection biases us away from detecting a negative relation between the variance of MPK and the level of diligence. The positive bias occurs because the variance of MPK declines with the level of diligence, and hence extreme values of MPK are truncated more often at lower diligence levels. Truncation therefore shifts the best-fit down at low levels of diligence, resulting in a slope coefficient that is spuriously high.

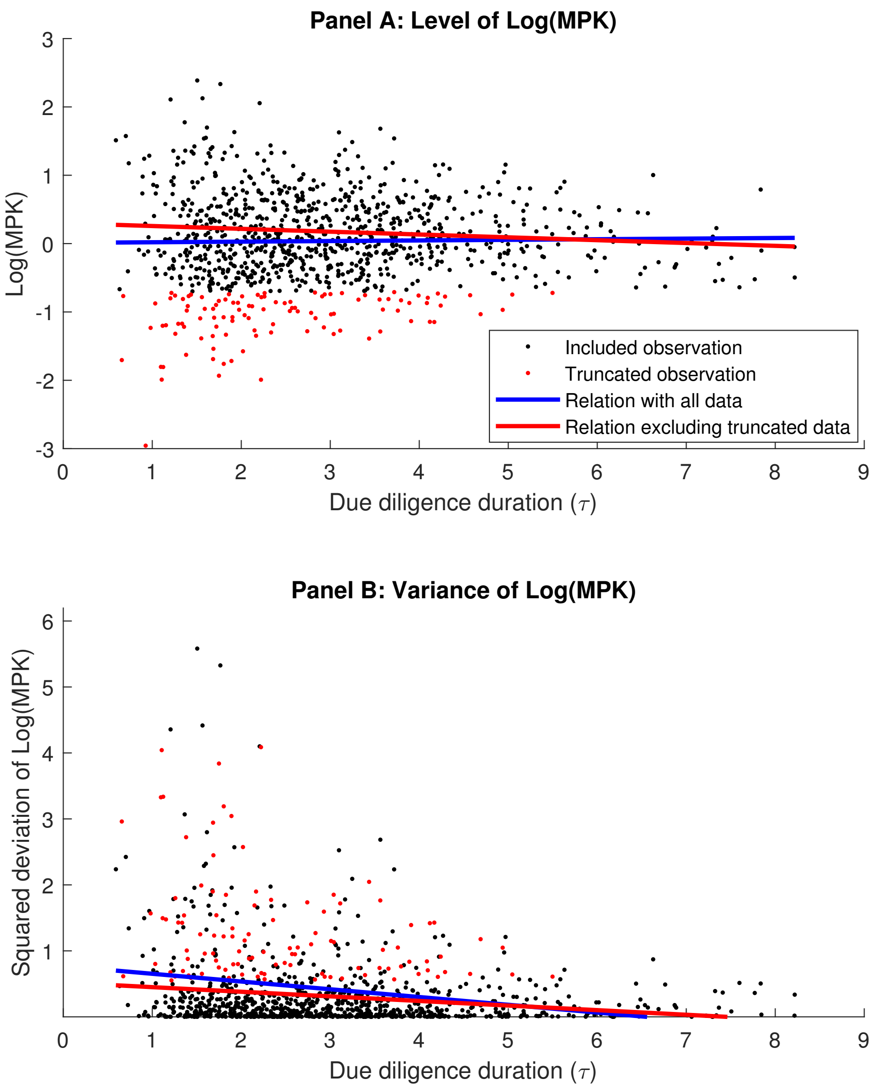

Figure B.1. Selection bias in the relation between due diligence and investment outcomes. We simulate data from the model in Section 5 using parameter values $\mu_0 = 0$ and $\nu_0 = 0.1$. We mechanically simulate a distribution of τ, which can be interpreted as resulting from a distribution of the diligence cost parameter c. We consider the case in which values of $\log(MPK)$ below -0.7 are truncated; those observations are plotted as red dots. The blue lines are the best-fit lines through all data points (included and truncated). The red lines are the best-fit lines through only the included observations. For ease of viewing, we limit the y-axis range in Panel B to exclude a few positive outliers.

B.3. Additional empirical results

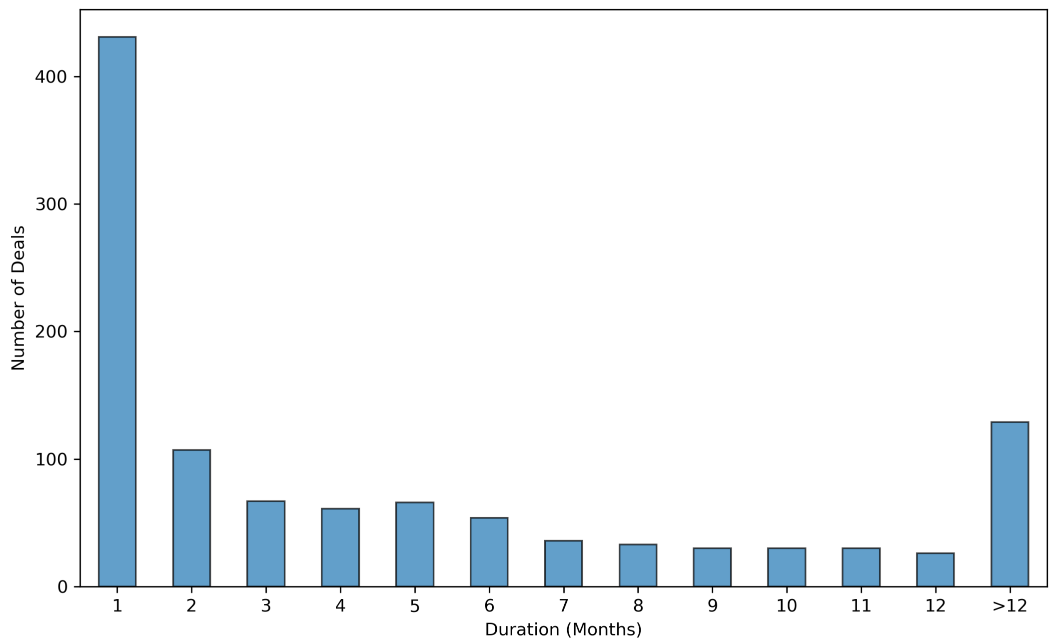

Figure B.2. Distribution of duration from pitch to investment. This figure shows the distribution of the duration from the pitch date to the investment date for deals with at least one captured pitch session within 18 months before the investment date. For deals with multiple pitch sessions, the last pitch session is used to compute the duration. Only deals between the lead investor and the startup are included. The average duration is 4.73 months, with the 25th, 50th (median), and 75th percentiles being 1, 2, and 7 months, respectively. Overall, 56.5% of the sample has a pitch-to-investment duration within 3 months, which is similar to the findings in First Round Capital's 2017 "State of Startups" survey of 869 venture-backed startup founders, where 52.7% reported a fundraising process duration within 3 months.

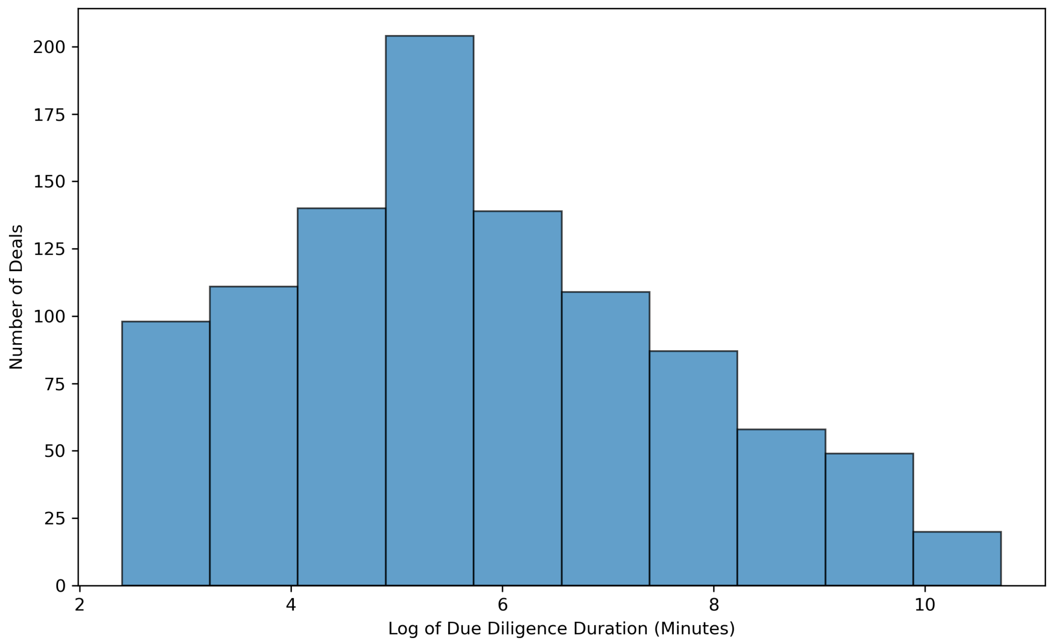

Figure B.3. Distribution of the due diligence measure. This figure plots the distribution of the log due diligence measure for deals with positive values. The mean due diligence duration in this subsample corresponds to 5.4 hours, while the 25th, 50th (median), and 75th percentiles correspond to 1.3, 4.5, and 19.3 hours, respectively.

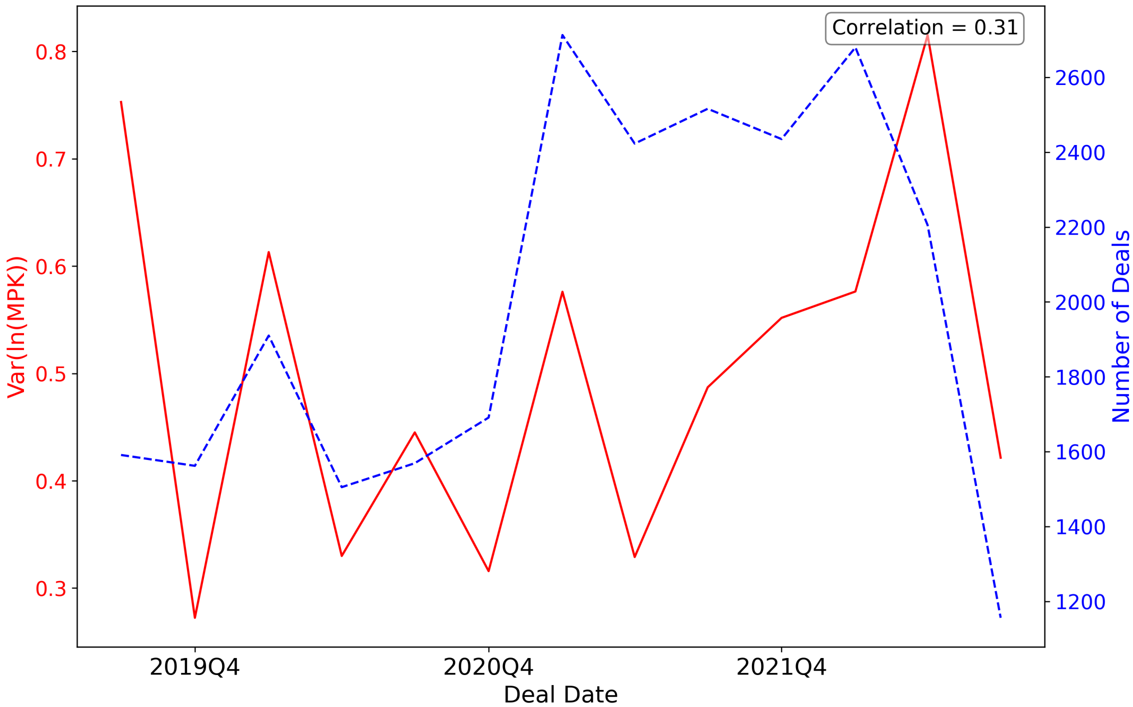

Figure B.4. MPK dispersion and aggregate VC deal volume. The solid red line (left y-axis) shows the median $Var(\log(MPK))$ across deals completed in each quarter. This measure follows the method described in the caption of Table 5, Panel A. To avoid data truncation issues, the first 18 months are excluded. The dashed blue line (right y-axis) shows the total number of deals categorized as "Early Stage VC" or "Later Stage VC" completed in each quarter. The correlation between the two series is 0.31, which is not statistically different from zero at the 10% level.

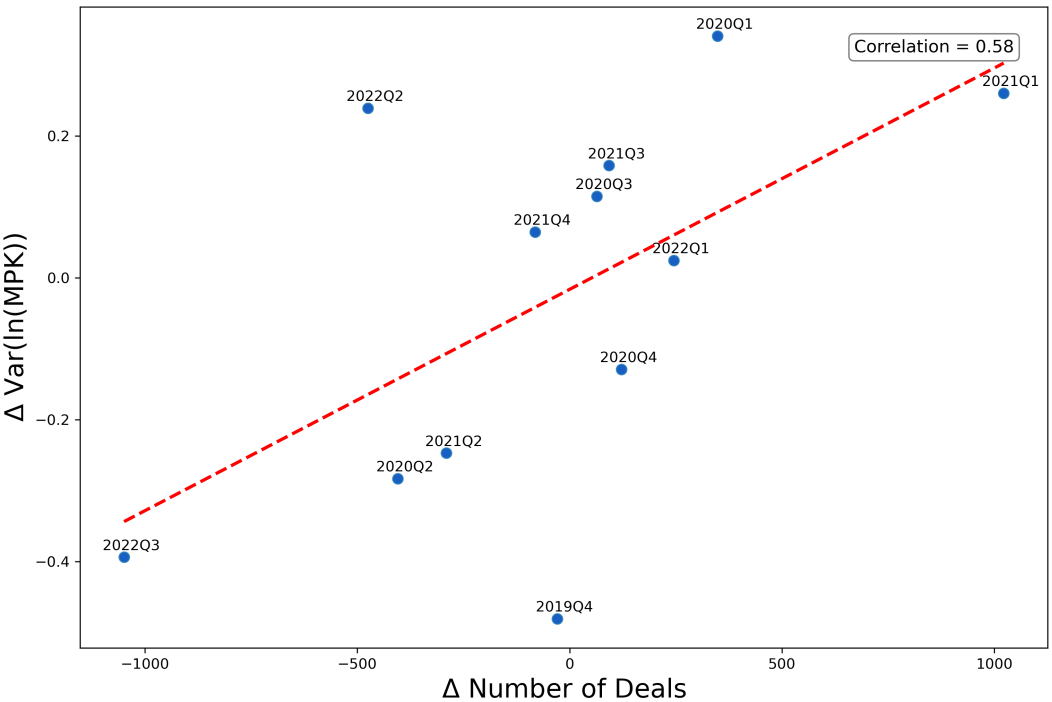

Figure B.5. First differences in MPK dispersion and aggregate VC deal volume. This figure shows the relation between the first difference in $Var(\log(MPK))$ and aggregate VC deal volume across quarters. The x-axis represents the change in the number of VC deals from the previous quarter to the current quarter. The y-axis shows the change in the median deal's $Var(\log(MPK))$ from the previous quarter to the current quarter. Each point is labeled with the corresponding quarter, and the fitted trend line is displayed as a red dashed line. The correlation is 0.58, which is statistically significant at the 1% level. This figure is constructed similarly to Figure 5. One difference is that we do not shift the series by one quarter. In Figure 5, we apply a one-quarter shift because due diligence typically occurs in the quarter prior to the investment. In contrast, both misallocation and the number of deals reflect conditions for the same quarter (i.e., for each quarter, how many deals were invested and how efficient they were). Therefore, comparing them without a shift is more appropriate.

Table B.1
Robustness: Alternative due diligence measure

This table presents robustness checks using an alternative DD measure: the number of months between the first meeting and deal date, considering only meetings within an 18-month pre-deal window. Observations with no meetings in this window have DD set to 0. Panel A replicates Table 4 using the log of one plus DD as the dependent variable, restricted to positive DD observations. Panel B replicates Table 5 using $Var(\ln(MPK))$ as the dependent variable. This table and all other tables in the Online Appendix follow their counterparts in the main text by using the same fixed effects and clustering by industry. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
Panel A: Dependent variable log(DD months)					
log(VC Contacts per Month)	-0.067	-0.082*	-0.086*	-0.074**	-0.065
	(0.038)	(0.037)	(0.037)	(0.030)	(0.037)
log(Distance)	-0.067***	-0.110***	-0.108***	-0.109***	-0.108***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.010)
Observations	766	765	765	765	682
Adjusted R^2	0.025	0.221	0.220	0.220	0.253
Panel B: Dependent variable Var(ln(MPK))					
log(DD months)	-0.240**	-0.191**	-0.191**	-0.114	-0.066
	(0.069)	(0.062)	(0.062)	(0.065)	(0.084)
Observations	5206	5201	5201	5201	5148
Adjusted R^2	0.001	0.020	0.020	0.169	0.164
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Table B.2
Robustness: Including non-lead investors

This table presents robustness checks including due diligence from all co-investors in a given funding round. Panel A replicates Table 4 using observations at the investor-startup level, as variables like distance are investor-startup specific. Since we continue to cluster by industry, the multiple observations in the same deal will automatically be clustered together. Panel B replicates Table 5. Since the dependent variable Var(ln(MPK)) is defined at the deal level, we aggregate observations to the deal level, with DD representing the sum of DD hours from all co-investors. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
Panel A: Dependent variable log(DD), including non-lead investors					
log(VC Contacts per Month)	-0.172***	-0.163***	-0.152***	-0.141***	-0.116**
	(0.030)	(0.030)	(0.029)	(0.028)	(0.032)
log(Distance)	-0.500***	-0.504***	-0.504***	-0.504***	-0.497***
	(0.027)	(0.026)	(0.026)	(0.027)	(0.020)
Observations	2786	2786	2786	2786	2747
Adjusted R^2	0.169	0.173	0.173	0.172	0.172
Panel B: Dependent variable Var(ln(MPK)), including non-lead investors					
log(DD)	-0.071*	-0.062*	-0.063*	-0.039	-0.033
	(0.029)	(0.029)	(0.028)	(0.024)	(0.022)
Observations	8753	8749	8749	8749	8706
Adjusted R^2	0.001	0.013	0.014	0.120	0.113
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Table B.3

Robustness: Excluding large VCs

This table presents robustness checks by excluding large VCs, which are more likely to have multiple office addresses. Using PitchBook AUM data, we exclude VCs above the 90th and 80th percentiles. Panel A (Panel B) replicates Table 4 with log(DD) as the dependent variable, excluding VCs above the 90th (80th) percentile. Panel C (Panel D) replicates Table 5 with $Var(\log(MPK))$ as the dependent variable, excluding VCs above the 90th (80th) percentile. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
Panel A: Dependent variable log(DD), excluding top 10% large VCs					
log(VC Contacts per Month)	-0.215*	-0.237**	-0.225*	-0.203*	-0.196
	(0.098)	(0.092)	(0.100)	(0.097)	(0.129)
log(Distance)	-0.662***	-0.690***	-0.688***	-0.694***	-0.731***
	(0.030)	(0.023)	(0.026)	(0.024)	(0.015)
Observations	701	700	700	700	614
Adjusted R^2	0.224	0.233	0.236	0.237	0.266
Panel B: Dependent variable log(DD), excluding top 20% large VCs					
log(VC Contacts per Month)	-0.204*	-0.231**	-0.225**	-0.209*	-0.233*
	(0.086)	(0.079)	(0.089)	(0.088)	(0.106)
log(Distance)	-0.663***	-0.697***	-0.695***	-0.701***	-0.728***
	(0.051)	(0.043)	(0.044)	(0.038)	(0.042)
Observations	638	637	637	637	549
Adjusted R^2	0.229	0.240	0.240	0.241	0.272
Panel C: Dependent variable Var(ln(MPK)), excluding top 10% large VCs					
log(DD)	-0.082***	-0.065***	-0.065***	-0.043**	-0.029
	(0.021)	(0.016)	(0.015)	(0.015)	(0.019)
Observations	4080	4075	4075	4075	4016
Adjusted R^2	0.001	0.023	0.023	0.170	0.163
Panel D: Dependent variable Var(ln(MPK)), excluding top 20% large VCs					
log(DD)	-0.063*	-0.062**	-0.062**	-0.065***	-0.047**
	(0.027)	(0.018)	(0.018)	(0.016)	(0.015)
Observations	3223	3219	3219	3219	3163
Adjusted R^2	0.000	0.018	0.019	0.158	0.144
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Table B.4

Robustness: Excluding same-building companies

This table presents robustness checks after dropping observations where VCs or startups share the same building address (i.e., identical street address, excluding suite numbers) as any other PitchBook-listed company. Panels A and B replicate Table 4 using log(DD) as the dependent variable. Panel A excludes startups co-located with other companies; Panel B excludes VCs co-located with other companies. Panels C and D replicate Table 5 using Var(ln(MPK)) as the dependent variable: Panel C excludes startups co-located with other companies; Panel D excludes VCs co-located with other companies. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
Panel A: Dependent variable log(DD), excluding co-located startups					
log(VC Contacts per Month)	-0.311**	-0.336**	-0.341**	-0.298**	-0.362***
	(0.096)	(0.102)	(0.106)	(0.119)	(0.071)
log(Distance)	-0.699***	-0.754***	-0.749***	-0.752***	-0.787***
	(0.086)	(0.097)	(0.098)	(0.086)	(0.122)
Observations	381	381	381	381	294
Adjusted R^2	0.225	0.248	0.239	0.244	0.258
Panel B: Dependent variable log(DD), excluding co-located VCs					
log(VC Contacts per Month)	-0.319	-0.363	-0.365	-0.405**	-0.639**
	(0.170)	(0.198)	(0.210)	(0.132)	(0.157)
log(Distance)	-0.745***	-0.793***	-0.837***	-0.944***	-0.887***
	(0.057)	(0.071)	(0.075)	(0.075)	(0.142)
Observations	122	104	103	102	61
Adjusted R^2	0.259	0.267	0.265	0.297	0.181
Panel C: Dependent variable Var(ln(MPK)), excluding co-located startups					
log(DD)	-0.085*	-0.053	-0.053	-0.039**	-0.051***
	(0.035)	(0.031)	(0.033)	(0.014)	(0.013)
Observations	1703	1702	1702	1702	1635
Adjusted R^2	0.000	0.015	0.014	0.160	0.154
Panel D: Dependent variable Var(ln(MPK)), excluding co-located VCs					
log(DD)	-0.111***	-0.080**	-0.104**	-0.109***	-0.124***
	(0.012)	(0.030)	(0.028)	(0.020)	(0.008)
Observations	574	573	573	573	493
Adjusted R^2	0.001	0.011	0.010	0.008	-0.072
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Table B.5

Robustness: Imputing zeros for missing next-round valuations

This table shows robustness to imputing zeros for missing next-round valuations. In Table 5, MPK is treated as missing in deals without subsequent-round valuation data in PitchBook. Here, for deals before 2020 (as 90% of follow-on rounds occur within 3 years), we impute zeros for missing subsequent valuations. Since a zero next-round valuation makes MPK negative and thus $\log(MPK)$ undefined, we try modeling $Var(MPK)$ instead of $Var(\log(MPK))$. This change makes the dependent variable highly sensitive to outliers, so we model absolute rather than squared deviations in MPK, as follows. First, we regress deal i's MPK_i on $\log(DD_i)$ and store the fitted value as $\overline{MPK_i}$. Then, we compute the regression's dependent variable as $|MPK_i - \overline{MPK_i}|$. We deviate from the theory's predictions by not taking logs and by modeling absolute rather than squared deviations in MPK, but these tests still preserve the model's intuition that dispersion in MPK relates negatively to the level of due diligence. Panel A shows results without filling in missing next-round valuations, while still making the other changes above. Note in Panel A that, even before imputing zeros, our baseline results from Table 5 are robust to modeling absolute deviations in MPK. Panel B shows how the results change after imputing zeros for missing next-round valuations. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
Panel A: Dependent variable $\|MPK_i - \overline{MPK_i}\|$					
log(DD)	-3.225***	-2.752***	-2.747***	-2.569***	-2.441***
	(0.461)	(0.282)	(0.278)	(0.429)	(0.310)
Observations	5514	5509	5509	5509	5457
Adjusted R^2	0.000	-0.001	-0.001	0.024	-0.001
Panel B: Dependent variable $\|MPK_i - \overline{MPK_i}\|$, imputing zeros for missing valuations					
log(DD)	-1.658**	-1.410**	-1.451**	-1.021*	-1.073
	(0.463)	(0.457)	(0.483)	(0.468)	(0.596)
Observations	9457	9453	9453	9453	9410
Adjusted R^2	0.000	0.001	0.001	0.019	0.003
Month FE	No	Yes	Yes	Yes	No
Industry FE	No	No	Yes	Yes	No
Stage FE	No	No	No	Yes	Yes
Industry by Month FE	No	No	No	No	Yes

Table B.6
Robustness: Alternative data-cleaning methods

This table checks the robustness of the main regression results using different data-cleaning methods. The baseline result for comparison is column 4 in Panel A of Table 5, reflecting the main result under month, industry, and stage fixed effects. The general data-cleaning method is as follows: A VC employee's device is defined as one that appears near the VC building for at least 5 working days in a month and is observed for at least 2 months. If this device is detected near a startup's office and remains there for a while, it is considered a potential meeting. The baseline and alternative tests differ in the following filters: (1) The baseline requires the VC employee to stay for at least 10 minutes near the startup to count as a meeting, while tests 1 and 2 use 30 and 60 minutes, respectively. (2) The baseline requires the VC employee to stay for no more than 300 minutes near the startup to count as a meeting, while tests 3 and 4 use 180 and 600 minutes, respectively. (3) The baseline requires at least three signals observed during the meeting interval to be considered a meeting, while tests 5 and 6 set this requirement to 2 signals and 5 signals, respectively. (4) If multiple VC employees visit the startup building on the same day, the baseline uses the maximum duration as the meeting time, while tests 7 and 8 use the sum or average duration as the aggregation method. (5) In the baseline, if a VC visits a startup more than 10 days in a single month, it is considered a false positive and dropped from the data; tests 9 and 10 use 5 and 20 days as the filter, respectively. (6) The baseline focuses on interactions occurring within 18 months prior to the investment date, ignoring those beyond this timeframe, while tests 11 and 12 use 12- and 24-month windows, respectively. (7) The baseline considers both VC employees visiting startup buildings and startup employees visiting VC buildings, while tests 13 and 14 consider only meetings at either startup buildings or VC buildings, respectively. Coefficients are reported with standard errors clustered by industry, shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Tests	Coef	Std	Min	Max	Obs	Method	Days	Window	Building
Baseline	-0.034**	(0.013)	10	300	3	max	10	18	all
1	-0.040*	(0.017)	**30**	300	3	max	10	18	all
2	-0.034*	(0.016)	**60**	300	3	max	10	18	all
3	-0.026**	(0.010)	10	**180**	3	max	10	18	all
4	-0.031*	(0.013)	10	**600**	3	max	10	18	all
5	-0.035**	(0.011)	10	300	**2**	max	10	18	all
6	-0.036*	(0.016)	10	300	**5**	max	10	18	all
7	-0.035**	(0.012)	10	300	3	**sum**	10	18	all
8	-0.033*	(0.015)	10	300	3	**avg**	10	18	all
9	-0.059**	(0.018)	10	300	3	max	**5**	18	all
10	-0.029**	(0.009)	10	300	3	max	**20**	18	all
11	-0.035**	(0.011)	10	300	3	max	10	**12**	all
12	-0.030**	(0.010)	10	300	3	max	10	**24**	all
13	-0.033*	(0.016)	10	300	3	max	10	18	**startup**
14	-0.118***	(0.016)	10	300	3	max	10	18	**vc**

Table B.7
Robustness: Extensive margin

This table tests the robustness of the results in Table 9 by varying fixed effects, the top-VC and top-DD cutoffs, and outcome variables. The dependent variable in Columns 1–4 is "Raise Next Round," which equals one if the startup raised a subsequent financing round or exited after the current round. Columns 1 and 2 vary the inclusion of fixed effects: Column 1 includes no fixed effects; Column 2 includes stage and industry-by-month fixed effects. In the baseline (Table 9), the top-VC and top-DD cutoffs are 20%. Columns 3 and 4 change these cutoffs to 25% and 33%, respectively. Columns 5 and 6 change the outcome variable: Column 5 uses the log of invested capital in the current round (a proxy for startup quality in our model), and Column 6 uses the log of invested capital in the next round, conditional on having a subsequent round. Standard errors are clustered by industry. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Tests of adjacent-group differences (computed but not shown) yield the following: the difference between Group 2 and Group 1 is significant in Columns 2–6 and not significant in Column 1; the difference between Group 3 and Group 2 is significant in Columns 1, 2, 5, and 6 and not significant in Columns 3 and 4; the difference between Group 4 and Group 3 is significant in all columns; and the difference between Group 5 and Group 4 is not significant in any column.

	(1)	(2)	(3)	(4)	(5)	(6)
Group 1	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Group 2	0.009	0.018**	0.026***	0.022***	0.113***	0.077**
	(0.006)	(0.005)	(0.003)	(0.005)	(0.019)	(0.030)
Group 3	0.043**	0.049***	0.042**	0.033**	0.193***	0.310***
	(0.014)	(0.010)	(0.013)	(0.010)	(0.034)	(0.075)
Group 4	0.185***	0.138***	0.148***	0.143***	0.528***	0.761***
	(0.024)	(0.026)	(0.029)	(0.021)	(0.026)	(0.060)
Group 5	0.218***	0.185***	0.151***	0.146***	0.596***	0.780***
	(0.014)	(0.014)	(0.009)	(0.018)	(0.068)	(0.066)
Constant	0.318***	0.318***	0.316***	0.314***	1.800***	1.903***
	(0.010)	(0.001)	(0.001)	(0.001)	(0.003)	(0.005)
Observations	38330	38329	38751	40016	30447	20533
Adjusted R^2	0.003	0.168	0.168	0.170	0.534	0.320
Month FE	No	No	Yes	Yes	Yes	Yes
Industry FE	No	No	Yes	Yes	Yes	Yes
Stage FE	No	Yes	Yes	Yes	Yes	Yes
Industry by Month FE	No	Yes	No	No	No	No
Top VC Cutoff	20%	20%	25%	33%	20%	20%
Top DD Cutoff	20%	20%	25%	33%	20%	20%