Due Diligence and the Allocation of Venture Capital

Jack Xiaoyong Fu
Lucian A. Taylor*

November 27, 2024

Abstract

How do investors choose the intensity of their due diligence, and how does that choice affect investment outcomes? Using cell phone 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 and Taylor are at the Wharton School of the University of Pennsylvania. We are grateful for comments from seminar participants at the University of 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 academic research. Due diligence refers to the process of evaluating an asset prior to buying or investing in it. This process plays a major role in alternative asset classes such as private equity, VC, infrastructure, and real estate, and also in the merger and acquisition (M&A) process. Despite its practical importance, there is little empirical research on due diligence, likely due to a lack of data. We attempt to fill this gap in the literature by analyzing new data on investors' due diligence. We study two related questions: How do investors choose how much due diligence to perform, and how does that choice affect capital allocation and investment outcomes?

We study these questions in the context of VC investing, where due diligence plays an especially large role. In a recent survey, VCs report spending an average of 118 hours on due diligence per investment, making diligence a significant part of VCs' jobs (Gompers et al., 2020). Due diligence is a primary way that VCs select which startups to invest in. That selection, in turn, is ranked by VCs as the most important of three contributors to value creation, with the other two factors being deal flow and value-add (Gompers et al., 2020). In the VC setting, due diligence can have significant real effects, as it determines which startups receive financing and in what quantity. These capital allocation decisions influence which innovations are realized, which can ultimately impact economic growth.

We focus on one important activity that occurs during VC due diligence: in-person meetings between investors and startup employees in the months leading up to a financing round. We measure these meetings' total duration using data from cell phone signals around VC and startup office buildings. Using this proxy for due diligence intensity, we find that investors choose to perform less diligence when it is more costly: when the startup is farther away; when the market, sector, or deal is hotter; and when the investor is busier. These patterns align with a simple model in which investors choose how much to learn about a startup, subject to costs, before investing. In both the model and our data, less diligence is associated with more dispersed investment outcomes, as measured by the variance of investments' marginal products of capital (MPK). A similar pattern holds for VC returns. Intuitively, less due diligence implies more investor uncertainty, which can lead to extreme outcomes in either direction. By interpreting these facts through the lens of our model, we find that due diligence matters greatly for the efficient allocation of venture capital to startups.

To construct our proxy for the amount of due diligence on a VC deal, we analyze cell phone signals near VC and startup offices. This fully anonymous dataset includes cell phone location information with timestamps, recorded whenever an app is active or running in the background. While this technology fully complies with legal standards and contains no individually identifiable information, it still allows us to measure potential meetings between VCs and startup employees.

From this raw data, we construct our diligence proxy in two steps. First, we identify likely VC or startup employees based on the frequency and timing of cell phone signals near their respective office buildings. Second, we identify meetings from time spent by a VC employee near the startup's office or vice versa. We collect data on these meetings during the 18 months leading up to a known investment by a given VC in a given startup. The total number of hours spent in these meetings is our proxy for due diligence intensity. We compute this proxy for roughly 22,000 U.S. deals in the PitchBook dataset from 2018 to 2023. 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. With a slight abuse of labeling, we refer to all these investors as "VCs."

We predict that VCs perform less due diligence when it is more costly. We test this prediction using several proxies for the costs of due diligence. Several of these proxies relate to how "hot" the given deal, sector, or overall VC market is. We argue that due diligence is more costly 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 potential deals, 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: it increases the chance that the VC loses the deal to a competitor who is willing to invest faster, with less scrutiny. 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 measure the number of total VC deals per quarter.

Supporting our prediction, we find less due diligence when the deal, sector, or VC market is hotter. In the aggregate time series, VC deal volume and due diligence have correlations ranging from -33% to -85% across specifications. Despite our short time series, the correlations are statistically significant in most specifications, and they are especially strong in the post-COVID period, after 2020. In deal-level regressions, the level 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 intensity. Doubling the number of other VCs meeting with the startup is associated with a 13% decrease in diligence hours. These regression 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 number of recent deals to 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 due diligence. VCs in those cases may substitute toward virtual meetings, which our measure does not capture, but it is unclear whether virtual and in-person meetings are good substitutes.

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 marginal product of capital (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 the model's 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 find a similar negative relation between our diligence measure the variance of VC returns, which are highly correlated with investments' MPKs. We measure VC returns as the fraction increase in the startup's valuation between financing rounds.

According to the model, the variance of MPK captures the degree of capital misallocation that results from limited due diligence. This idea comes from Hsieh and Klenow (2009) and the very 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 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 due diligence improves the allocation of venture capital to startups. We use the model to quantify this effect. By feeding our estimated regression coefficients through the model, we measure how doubling the hours spent on due diligence changes 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 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. In some specifications, however, the relation is negative and statistically significant. The negative relation could result from VC risk aversion, which is outside our model. Facing greater uncertainty (resulting from less diligence), a risk-averse VC is inclined to under-invest, which generates higher performance (due to decreasing returns).

Beyond these main tests, we also provide interesting new descriptive facts about VC due diligence. Average diligence levels are similar between traditional VCs and accelerators/incubators, but diligence levels are lower for "nontraditionals" such as CVCs, growth/expansion investors, hedge funds, and other asset managers. On average, diligence levels are higher for investors who manage fewer assets and make fewer other deals in the surrounding months. Lastly, an investor's full-sample exit rate—a proxy for investment success—has a strong, positive correlation with its due diligence intensity.

Since our diligence measure is new, we perform several validation checks. If we look at 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 cell phone 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 validation 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 cell phone. While these measurement issues introduce noise, we argue that, in most cases, the measurement error is not correlated with the other variables we study, and hence the measurement problems do not bias in the direction of our conclusions. There is clearly more work to be done 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. Indirectly related to due diligence, several papers relate startup and deal characteristics to either investor interest or investment outcomes.[1]

There is little empirical research on due diligence even outside VC. Gompers et al. (2016) survey 79 private equity investors and describe their deal-selection process and criteria. Lajoux and Elson (2010), Cole, Ferris, and Melnik (2016), and Offenberg and Pirinsky (2015) describe the large costs and time spent on due diligence in M&A deals. Similar to us, Wangerin (2019) studies the causes and consequences of due diligence intensity, albeit

[1] See, e.g., Baum and Silverman (2004), Caliendo and Kritikos (2008), 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.

in an M&A setting. Wangerin measures the length of due diligence as the time between the acquisition agreement and transaction completion. Echoing one of our results, Wangerin finds that competitive pressures are associated with less due diligence. Despite the importance of due diligence in real estate (e.g., Brueggeman and Fisher, 2019), we are not aware of any studies on it. However, our finding of less diligence in hot VC markets echoes stories of buyers waiving home inspections in hot real estate markets (e.g., Ostrowski and Petry, 2023).

On the theory side, the idea that investors choose how much to learn about an asset before investing goes back at least to Grossman and Stiglitz (1980). Since then, a large literature has modeled learning by buyers. Most relevant to our study, Daley, Geelen, and Green (2024) model due diligence. They find that acquirers perform "too much" due diligence in equilibrium, and a simple quantification suggests the distortion is economically significant. Their paper and ours both model due diligence as learning about an asset under symmetric, imperfect information. Compared to our toy model, theirs is a rich framework that includes dynamics, deal pricing, and deal completion. Their theory is mostly tailored to the M&A setting, which differs from VC in notable ways. Specifically, in VC, due diligence occurs before the transaction price is finalized, and VCs inject new capital in exchange for a minority stake.

This paper also belongs to the large literature on capital misallocation, following Hsieh and Klenow (2009).[2] 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 builds on Fu (2024), who uses the same cell phone 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 cell phone geolocation data in economics is relatively new.[3]

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 validation tests 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 allocation.

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

[3]Other papers using cell phone geolocation data include Chen et al. (2022a, 2022b), Chen et al. (2023), and Atkin et al. (2024).

Section 6 empirically tests those implications and provides a simple quantification. Section 7 concludes.

2. Data and institutional background

2.1. Due diligence in VC

The goal of VC due diligence is to assess the 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 those 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 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

Our proxy for the intensity of VC due diligence is based on in-person meetings between VC and startup employees leading up to the investment. To create this measure, we analyze cell phone signals within a 200-meter radius around VC and startup office buildings. The sample period is from January 2018 to January 2023. The cell phone signal dataset is widely used across many industries to understand user behaviors. Smartphone operating systems (Android and iOS) record the longitude and latitude of a cell 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 in app stores, spanning a variety of categories, including messaging, social media, navigation, music, photo, weather, travel, health and fitness, and eight other categories.

The cell phone data coverage is quite good. The data provider reports coverage of 220 to 240 million monthly active users in the U.S., which represent roughly 80% of all smartphones.[4] For comparison, Testoni et al. (2022) use a geolocation dataset from November 2015 to November 2017, and they report coverage of only 10% of U.S. smartphone users.

To construct the dataset, we capture all cell phone signals near VC and startup office buildings. Using the Google Maps API, we calculate the distance between each signal and each office building's address. For the baseline regression, we use a 200-meter cutoff, discarding signals beyond this range while retaining those within it.

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 in the vicinity of the VC office for 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 by using the proximity and duration of a device's presence near a startup's office. 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

[4]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.

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 cell phone 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 timeframe. In the robustness tests (Table B.1 in the Online Appendix), we examine alternative parameters for each of these filters. 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 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 miss meetings if PitchBook lists the wrong address for a VC or startup. We also miss in-person meetings that occur in alternative locations, such as restaurants. Additionally, 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 cell phone usage habits (such as how often users clear background apps), is unlikely to correlate with specific VC or startup characteristics and therefore is not expected to introduce bias into the results. Our proxy also omits diligence activities that do not involve meetings between VC and startup employees. Examples include reference-checking, customer meetings, consulting external experts, and financial modeling. Our results clearly pertain to 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 intensity. To the extent that investors do not scale these activities proportionally, we measure total diligence intensity with error. Later, we discuss whether that error correlates with the variables we study.

2.3. Sample formation

We build our data sample using the following steps. 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 deals. We then gradually apply filters based on investor, startup, and deal characteristics. For investors, we apply the following filters: they must be 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 deals remain after applying these filters. Next, we require startups to have headquarters in the U.S. and a valid headquarter address (roughly 104,000 deals remaining). Last, we apply deal-specific filters. We require that the investor and startup have different office addresses (about 103,000 deals remaining). The deal type must be "Early Stage VC," "Later Stage VC," "Seed Round," or "Accelerator/Incubator" (about 91,000 deals remaining). Note that although the deal types are specific to the four categories above, to the extent that nontraditional investors—such as buyout firms and asset managers—also occasionally invest in VC deals, our sample includes nontraditional investors as well. We then require the investor to be the lead investor (about 26,000 deals 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 22,000 deals.

3. Stylized facts and validation

Table 1 contains stylized facts on the due diligence measure and other variables that we introduce later. Appendix Table 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 cell phone 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 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. Under the reasonable assumption that each meeting has a similar probability of appearing in our data, a deal with zero recorded meetings likely has fewer total meetings than a deal with a positive number of recorded meetings.

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. The measure has a roughly lognormal distribution (see Figure B.2 in the Online Appendix).

As a validation check, we compare our summary statistics to those in Gompers et al. (2020), which contains results from 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 intensity across different subsamples. Panel A shows strikingly 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 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. We can reconcile these results if a decreasing share of total due diligence hours are spent in in-person meetings as companies mature: total hours can increase with maturity while hours spent in in-person meetings (i.e., the hours we measure) stay flat. Lending plausibility to this story, at very early stages, relatively little diligence can be done beyond in-person meetings; there are often no customers or product yet to analyze, and it is usually too early to perform complex financial modeling or data analysis. Instead, at early stages, investors focus more on the team (Gompers et al., 2020), and evaluating the team likely requires in-person meetings. Given these important differences in the nature of due diligence across stages, we include stage fixed effects in our main tests.

Panel B compares 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). We also find more diligence in the materials and resources industry, which makes up a small part of the sample. 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 check 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.[5]

For further validation, we use our cell phone 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.[6] 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 cell phone 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

[5]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).

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

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, 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.[7] 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. We find that in 32% of rounds, the startup pitched to more than 20 investors, compared to 23% in the survey. These numbers are all quite close, again suggesting that our method for using cell phone data to identify startup-investor meetings is reasonable.

Figure 3 relates due diligence intensity to investor characteristics. 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 VCs firms do less due diligence on average (Panel A), perhaps because their stronger reputations attract higher-quality deals, which require less scrutiny.

Panel B shows a strong positive relation between diligence levels and the investor's exit rate, a measure of investor success. This exit rate is measured at the end of the sample period and is held constant across all the investor's deals. One possible story is that due diligence improves investment outcomes, a hypothesis we study more carefully later.

In Panel C, 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, Nanda, and Rhodes-Kropf, 2018), making many investments with little due diligence. Other VCs follow a more selective strategy, making fewer investments but performing more diligence. Another potential explanation is that VCs do more deals and perform less diligence when markets are hotter, as discussed above. In either explanation, investors with more recent deals are busier, making due diligence more costly.

A confounding factor is that VC firms with many deals can spread the work across more employees. To better measure how busy investors are, Panel D considers the number of recent deals per investment professional at the VC firm. We still see a negative relation, consistent with busier investors doing less diligence, although the relation is largely driven

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

by the far right tail of busyness. We explore this pattern more carefully in the next section.

4. The due diligence choice

Whether to perform more or less due diligence is a choice made by the investors. That choice is influenced by the 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 overall is "hotter," i.e., when more deals are being made. Due diligence becomes more costly in hot markets for three potentials reasons. First, in hot markets there is a higher risk of having a deal picked off by another investor during the diligence period.[8] Second, the opportunity cost of a VC's time is arguably 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.

At the VC industry level, there is indeed a negative relation between due diligence intensity 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 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.[9]

[8]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.

[9]According 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).

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 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.[10] For example, if earlier-stage companies face higher uncertainty and therefore require more diligence, or if in-person diligence is relatively more 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.[11] One potential concern about this result is that investors may substitute

[10]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.

[11]The estimated coefficient is -0.632, so doubling Distance changes the diligence measure by a fraction

virtual meetings for in-person meetings when distances are large, and virtual meetings are missing from our data. If virtual and in-person meetings were perfect substitutes, then our results would not necessarily imply less total diligence—including unmeasured diligence—when distances are larger. However, to the extent they are not perfect substitutes, our result points to a negative relation between diligence and distance.

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 within a specific time window, defined as the shorter of either 18 months or the period between the focal VC's first pitch date and the investment date. When calculating the number of other VCs, we exclude the focal VC and all existing 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 of 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 at the 5% level. Economic significance is quite 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.[12] The magnitude increases to a 34% reduction if we use the largest coefficient, from column 5.

$2^{-0.632} - 1 = -0.35$.

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

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 accountants and marketing directors. 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.

To summarize, we find a negative relation between the chosen length of due diligence and several proxies for 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. 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 of 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).

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 VC capital invested in the startup, and a, the unobservable quality of the VC-startup match. That match quality reflects both the startup's productivity 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 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 $t = 2$, 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 due diligence intensity and its costs.

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 marginal product of capital (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 more dispersion there is in MPK. As in Hsieh and Klenow (2009), in our model the variance of MPK provides a measure of capital misallocation. In our setting, this misallocation results from imperfect information. 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 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.

The model is deliberately simple and omits many features of reality. Chief among them, we assume in equation (1) that the VC maximizes the deal's total surplus, which of course is shared by the VC and the startup's existing owners. The VC's part of that surplus depends on deal terms, which are driven by competition and bargaining power—elements we omit. Daley et al. (2024) theoretically study the interplay of deal terms and due diligence. As an approximation of reality, we assume the VC first maximizes the total surplus through its choice of τ and K. Outside our model, the startup and VCs then split that surplus.

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 explains our variables' measurement. Section 6.2 tests Predictions 2 and 3 above, which are about the variance and level of MPK, respectively. 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 or lost through due diligence.

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.[13] 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 valuations for rounds, 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

[13] Even if θ varies across deals, our regression coefficients are not necessarily biased. Our dependent variable of interest is the variance of our log MPK proxy. The regression residuals therefore include the variance in $\log(\theta)$. As long as the variance in $\log(\theta)$ is not related to our due diligence measure, then the coefficients of interest are not biased.

valuation.[14] 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 $Var(MPK)$ upward. Measurement error will not necessarily bias our results, since the MPK measures are on our regressions' left-hand sides, and it is unlikely that the unexpected ex-post shocks are correlated with our main regressors, the due diligence measures. The upward bias in $Var(MPK)$ is also not necessarily problem, as we are not trying to measure the absolute level of this variance. Instead, our goal is to relate $Var(MPK)$ to our due diligence measures. As long as the shocks' volatilities are 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. First, 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.[15]

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. Second, when our dependent variable is $Var(MPK)$, 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. Finally, 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 data on pre-money valuations are often missing from PitchBook. This challenge reduces our sample size but does not obviously bias our results. While sample selection issues arguably bias the average level of valuation step-ups, that is not the object we study. Instead, we are interested in the relation between due

[14] 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.' If a deal lacks information on a subsequent round in PitchBook, we treat MPK as missing in our baseline analysis. In a robustness test, for deals made before [2023 - τ] without subsequent round information, we set the subsequent round valuation to zero. We set τ to three years, as over 90% of deals in our sample raised the next round within this time frame. The robustness test yields similar results.

[15] We set r to the industry's stock return between the current and next financing round. 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. Our approach implicitly assumes the startup has a beta of 1 on its industry index.

diligence measures and investment outcomes, and it is not obvious that sample selection would bias those correlations.

Third, complex VC deal structures can make pre-money valuations poor approximations of companies' markets prices (Gornall and Strebulaev, 2020, Metrick and Yasuda, 2021). This is yet another source of measurement error. Mitigating this concern, the measurement error in $Var(MPK)$ 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 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 $Var(\log(MPK_i)|X_i)$.[16] Since we have the variable's realized rather than expected value, we measure variance with error, but this error is on our regressions' left-hand side. We model the variance of returns similarly.

6.2. Reduced-form evidence

Table 5 shows our tests of Prediction 2. 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

[16] 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.

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 a one standard deviation increase in log due diligence is associated with a 0.12 decline in the variance of $\log(MPK)$.[17] 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%.[18] This decline appears modest primarily because ex-post shocks make the measured dispersion in MPK very high. If we could instead 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. Those 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 has little effect on the due diligence coefficents. Their magnitudes increase, but only slightly. The controls soak up residual variance, which increases the due diligence coefficients' statistical significance. We control for the log of startup age as a more direct proxy for prior uncertainty. Consistent with Prediction 2, startup age enters with a negative coefficient, but only in the presence of stage FEs. 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. Finally, we control for the duration between this round and the next in order to soak up variation in ex-post shock

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

[18]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%.

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 diligence. To include those observations while still taking logs, we use the transformation $\log(1+x)$.[19] 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.[20] 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, we find a significant negative relation between diligence intensity and dispersion in MPK in almost all specifications. Interpreted through our model, the result indicates that less due diligence leads VCs to invest under greater uncertainty, which leads to more over- and under-investment—that is, more capital misallocation. In both our model and the regressions, diligence intensity 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. 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

[19] Chen and Roth (2024) show several problems with the $\log(1+Y)$ transformation for outcome variables Y. We apply the transformation to a regressor, not an outcome variable.

[20] 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.

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. A potential explanation for a negative relation could be that VCs' risk aversion leads them to under-invest when they face more uncertainty (resulting from less diligence). We never find a positive relation, meaning we find no evidence supporting the notion that VCs often "throw money at deals," performing little due diligence and over-investing, which would produce bad performance.

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 to 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 positive-diligence sample. 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.[21] Table 8 relates the due diligence measure to the level of return. That relation is statistically insignificant in 11 out of 15 specifications. The slope coefficients are negative in all but one specification. Similar to before, we find mixed evidence on the relation between due diligence and the level of performance, with a few specifications pointing to a negative relation.

[21] 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%.

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 an increase in due diligence improves the value created by VC investments.

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

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

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 this effect, 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 startup's true 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[22].

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 reduces the average value lost from imperfect information by 6% of the VC's amount invested.[23]

[22]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.

[23]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%×2M.

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 created 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 $1,525 of extra value per hour of extra work. This number seems plausible, in the sense that the opportunity cost of a VC's time might exceed this amount, in which case the extra effort on due diligence might not be privately optimal. Even if the VC's time is worth less than $1,525 per hour, spending extra time on due diligence might not be privately optimal, because doing so could allow the deal to be stolen by a competing, faster VC.

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 to value creation changes from 6% to 17%, and it increases further if we use a higher values of θ. On the low end, the improvement to 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.

7. Conclusion

This paper provides new empirical insights into how VCs adjust the intensity of their due diligence in response to various factors, such as market conditions and investor workloads.

By using cell phone signal data to measure the duration of pre-investment meetings, we show that VCs perform less due diligence when it is more costly—when travel distance is greater, when VCs are busier, or when the deal, sector, or overall VC market is hotter.

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 substantial. Taken together, our results indicate that investors deliberately choose the intensity of due diligence, weighing its costs against the improvements to capital allocation.

The due diligence process 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. We hope future research will refine the measurement of due diligence, not only in VC but also in private equity, real estate, M&A, and other areas where it 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.

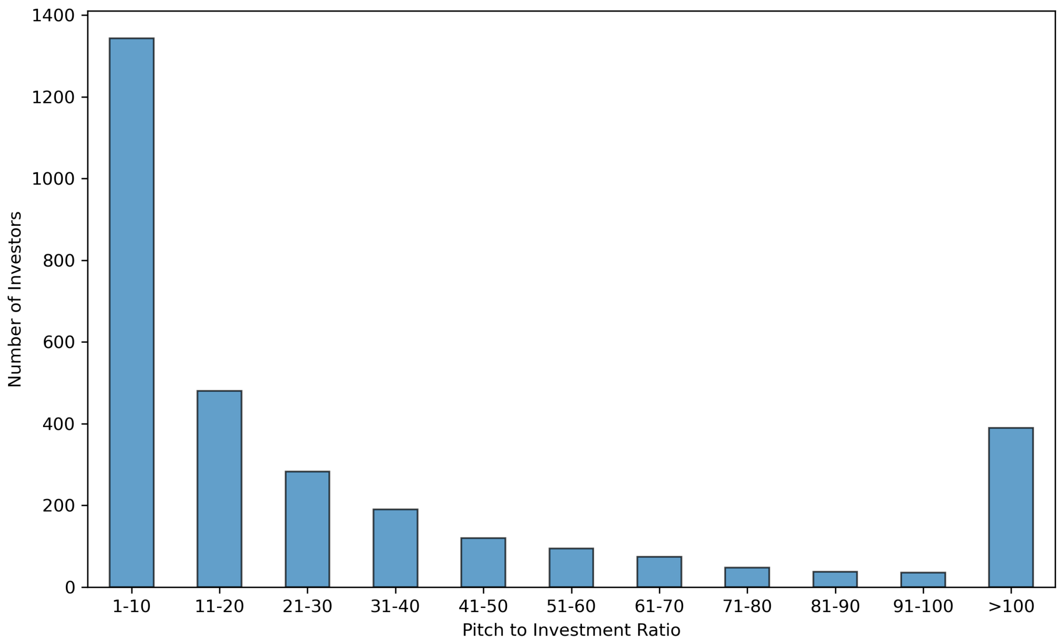

Figure 1. Number of pitches VCs receive per investment. This figure shows the distribution of the pitch-to-investment ratio for investors. The ratio is calculated as the number of captured pitches between a focal investor and any startup from 2018 to 2023, divided by the total number of investments made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

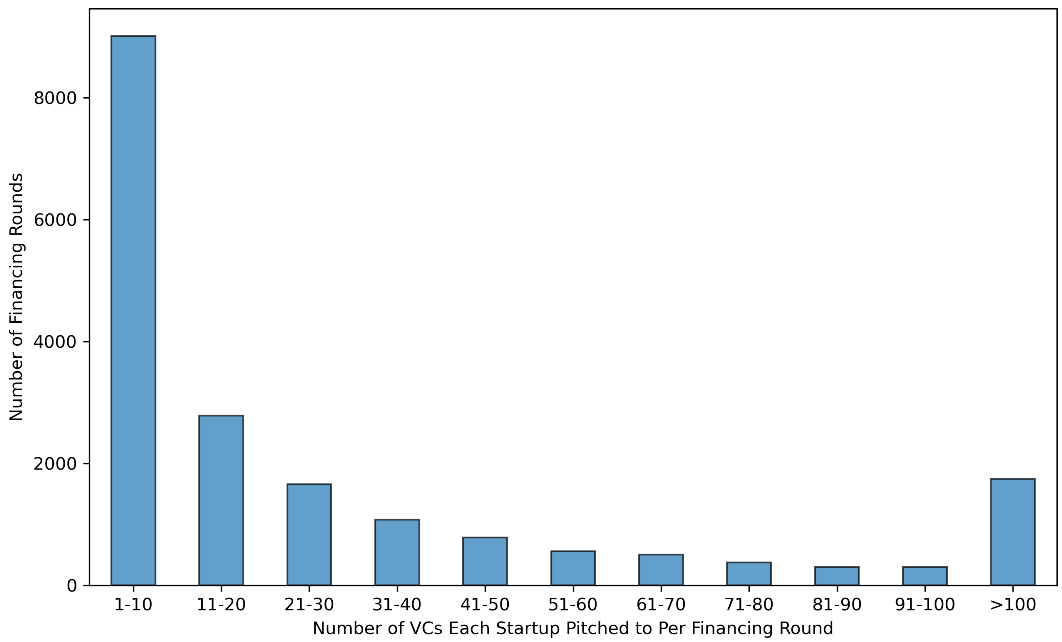

Figure 2. Number of pitches startups make per funding round. This figure shows the distribution of the number of investors that a startup pitches to per funding round during the sample period from 2018 to 2023. Each financing round is treated as an individual observation. The average number is 27, with 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.

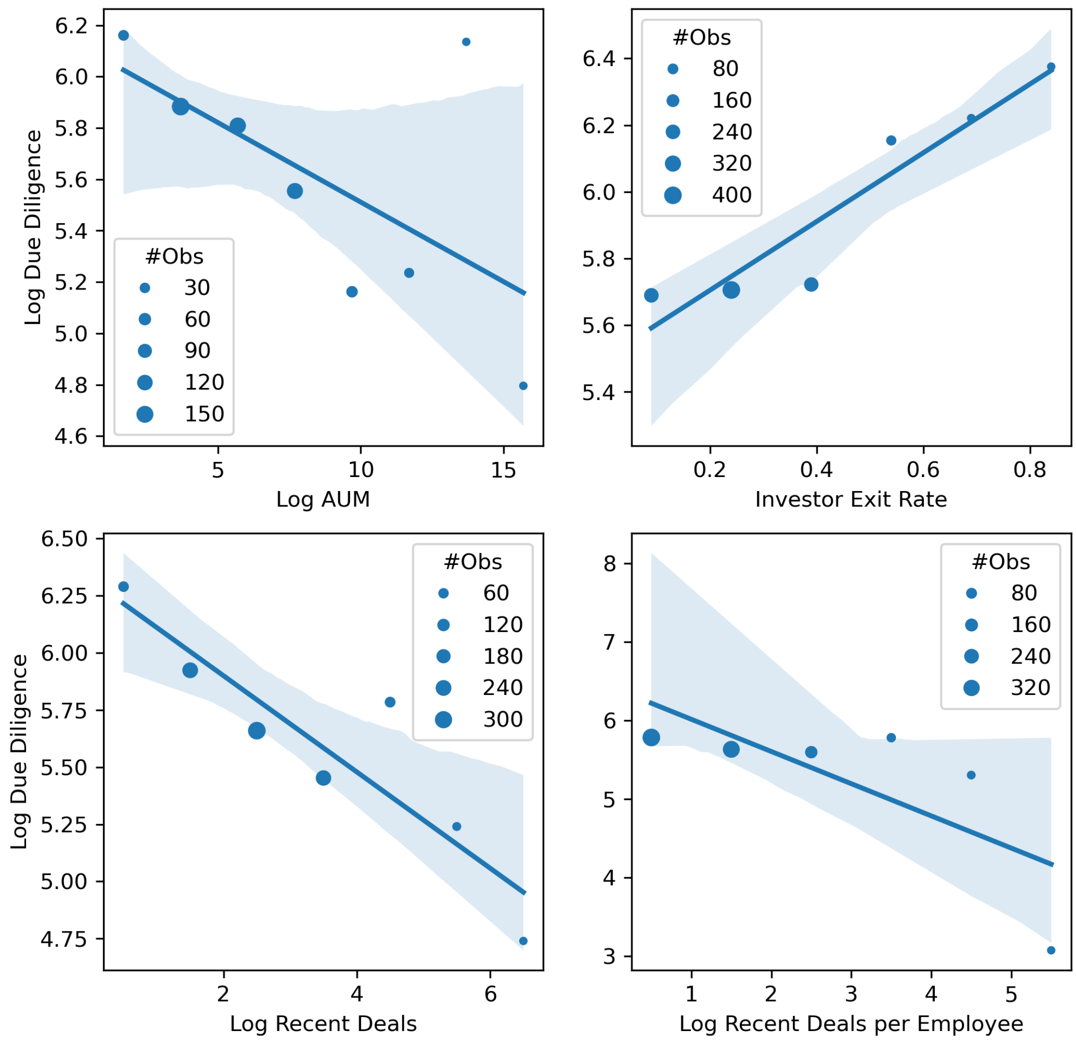

Figure 3. Due diligence and investor characteristics. This figure shows the relation between due diligence duration and investor characteristics across four subplots. In all four graphs, each observation represents an individual deal, rather than an investor. The top-left subplot shows the log of the PitchBook variable 'AUM' on the x-axis, measured in millions of dollars for the investor associated with the focal deal. The top-right subplot displays the exit rate, defined as the ratio of the PitchBook variable 'Total Exits' to the PitchBook variable 'Total Investments' for the focal investor. For both AUM and exit rate, we use the most recent values available in PitchBook, and we keep them constant across different deal dates. The bottom-left 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's investment date. The

bottom-right subplot presents the log of the ratio of number of recent deals to 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. These two variables depend on both the investor and the deal date. The y-axis for all subplots is the log of due diligence duration, measured in minutes, where 3 stands for approximately 0.8 hours, 5 stands for approximately 2.7 hours, and 7 stands for approximately 8.2 hours. All samples represent interactions between the lead investor and the startup, considering only new investors in the focal round. Only deals with positive due diligence duration are used to plot the figures. The bands around the regression line show the 95% confidence intervals, estimated using a bootstrap.

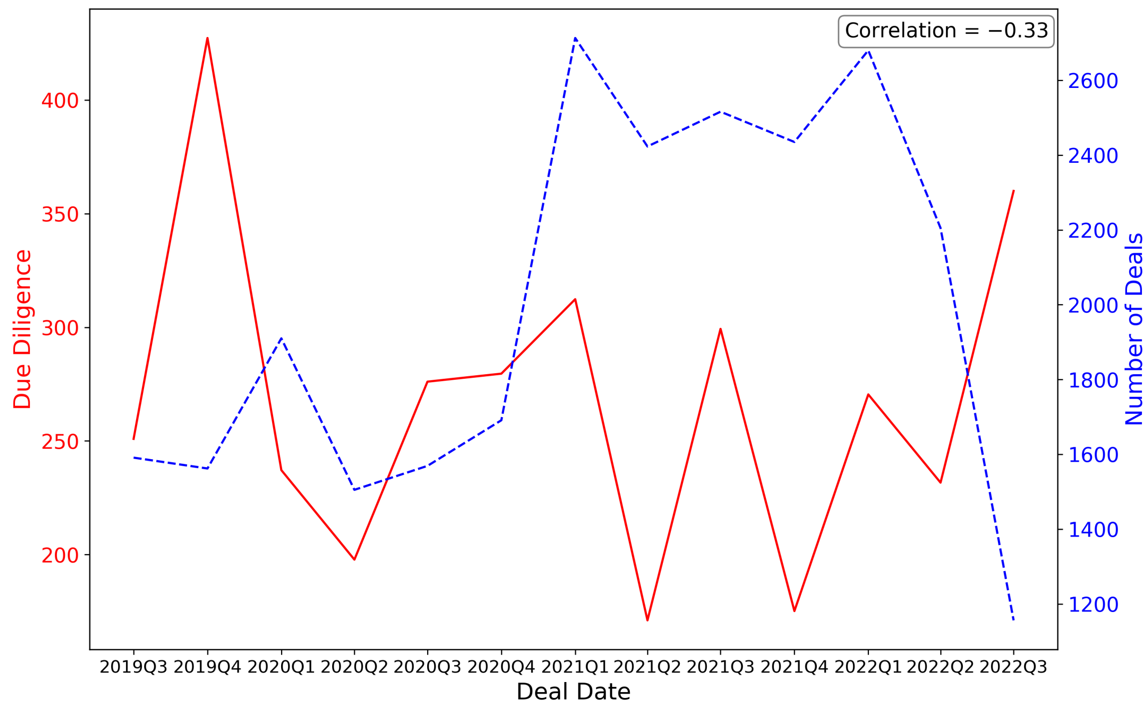

Figure 4. Due diligence and deal volume. This figure shows the due diligence duration and deal volume over time. The solid red line (left y-axis) represents the median due diligence duration for deals invested in each quarter. Due diligence duration is measured as the total number of minutes that an 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.

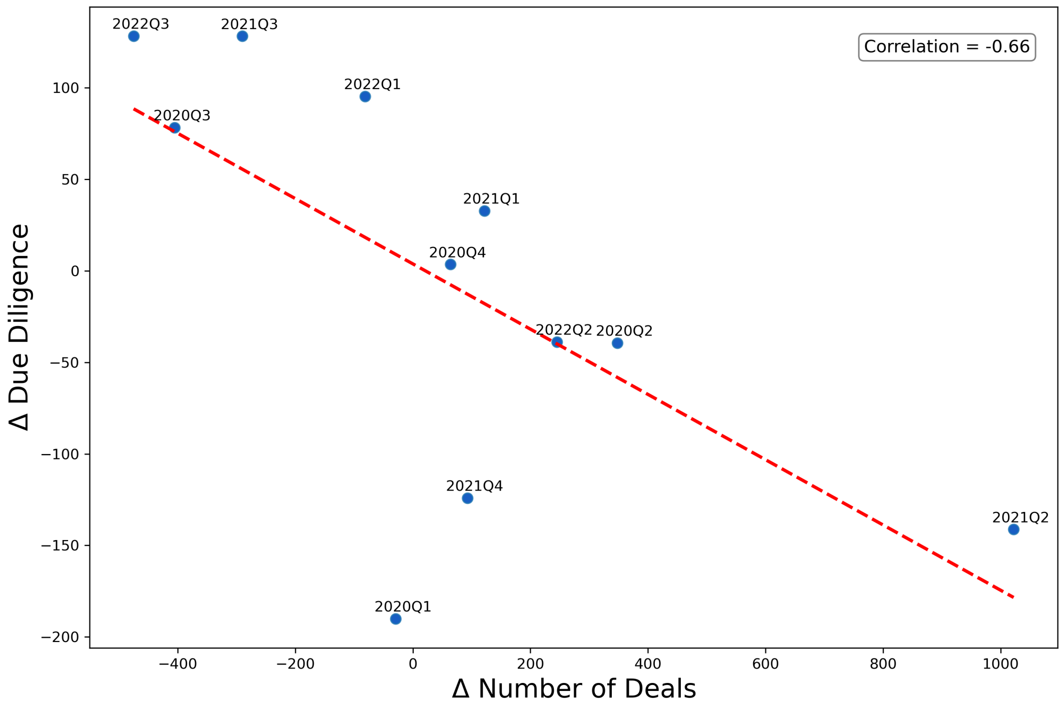

Figure 5. Quarterly changes in due diligence and deal volume. This figure illustrates the relationship between the first difference in due diligence duration and deal volume across quarters. The x-axis represents the change in the number of VC deals from the previous quarter to the current quarter, shifted 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.

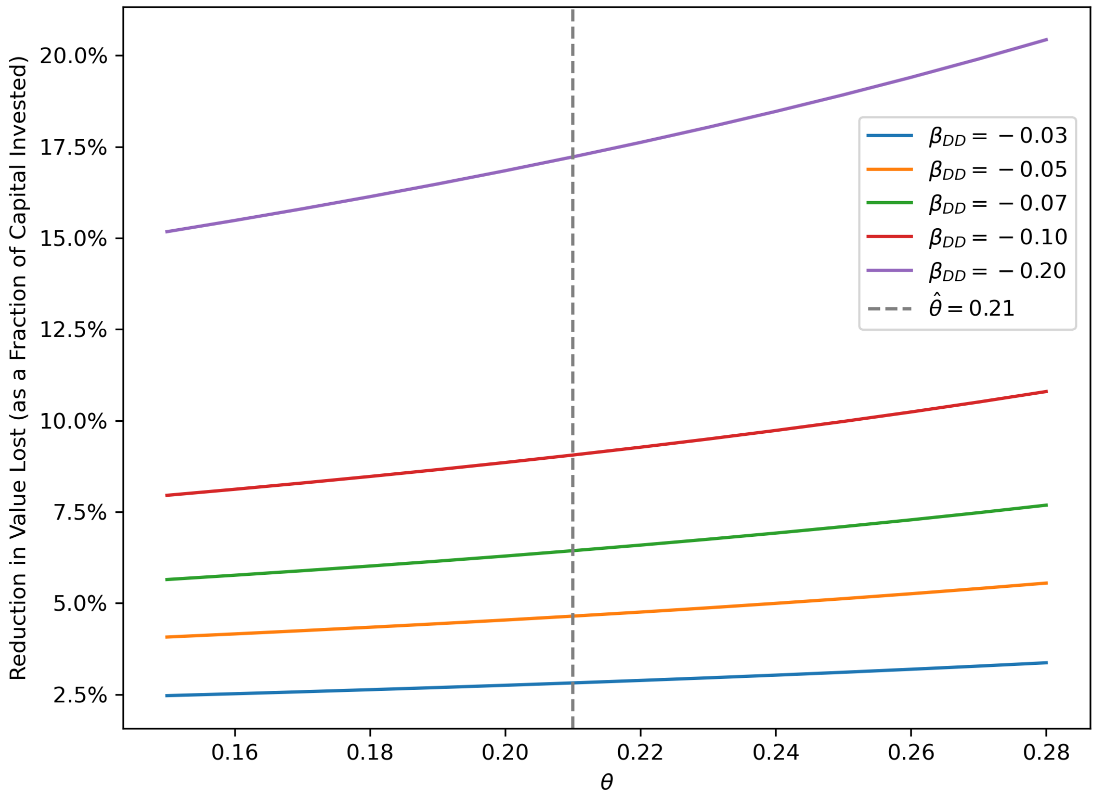

Figure 6. Due diligence and improvements to capital allocation. This figure illustrates how much the value lost due to imperfect information can be reduced by doubling due diligence hours, based on various regression coefficients from Table 5 and different values of the curvature parameter θ. The x-axis represents θ, ranging from 0.15 to 0.28, which corresponds to the 95% confidence interval for $\hat{\theta}$, which has a point estimate of 0.21. 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.

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 the data description section. Panel B includes only deals with positive due diligence duration. Due diligence is measured as the time that an investor and 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 of 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 in 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 due diligence duration, measured by the total hours investors and startups spent together within 18 months before the investment date, across 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: 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 the correlation coefficients between the time series of due diligence duration and VC deal volume. Due diligence duration is defined as the total minutes that an investor and startup spent together at either party's building within 18 months prior to the investment date. To avoid data truncation issues, the first 18 months of the sample period are excluded. To mitigate the impact of outliers, we use the median of positive due diligence samples. VC deal volume is measured as the total number of 'Early Stage VC' or 'Later Stage VC' deals invested in each quarter. Due diligence for a given quarter is represented by the median duration for 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 current 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 **, 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 due diligence duration, measured as the total number of minutes that the investor and startup spent together within 18 months prior to the investment date. All panels exclude observations with zero diligence duration. In Panel A, we use the baseline proxy for C, which is the distance between the investor's and startup's office buildings. The longer the distance, the higher the direct cost of conducting due diligence on-site. The proxy for competition is the log of Contact VC per Month, defined as 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. In Panel B, we use the same proxy for direct cost but use an alternative proxy for competition, which is Abnormal Deal Volume. This is computed as the deal volume for the same year, stage, and industry as the focal deal, divided by the average deal volume of the previous two years. In Panel C, we retain the same proxy for competition as in the baseline but change the proxy for C to the log of Deals per Partner. 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), $log(K)$, and Time to Next Round. Startup Age is the time from the startup's founding year to the deal date. 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)	-0.060***	-0.054***	-0.054***	-0.037***	-0.026
	(0.013)	(0.011)	(0.011)	(0.009)	(0.013)
log(Startup Age)	0.126**	0.132**	0.128**	-0.081	-0.109**
	(0.043)	(0.046)	(0.044)	(0.044)	(0.045)
log(K)	-0.434***	-0.435***	-0.429***	-0.437***	-0.434***
	(0.031)	(0.030)	(0.028)	(0.016)	(0.018)
Time to Next Round	0.025***	0.025***	0.026***	0.031***	0.032***
	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)
Observations	5168	5163	5163	5163	5111
Adjusted R^2	0.046	0.053	0.052	0.101	0.100
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 relationship 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)	-0.033**	-0.029**	-0.029**	-0.016	-0.014
	(0.010)	(0.008)	(0.008)	(0.009)	(0.009)
log(Startup Age)	-0.117**	-0.129**	-0.134**	-0.275***	-0.290***
	(0.038)	(0.040)	(0.044)	(0.037)	(0.034)
log(K)	-0.323***	-0.320***	-0.314***	-0.315***	-0.309***
	(0.027)	(0.025)	(0.023)	(0.021)	(0.021)
Time to Next Round	-0.024**	-0.020**	-0.020**	-0.016*	-0.017*
	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)
Observations	5168	5163	5163	5163	5111
Adjusted R^2	0.158	0.188	0.193	0.308	0.324
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 relationship 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)	-0.031***	-0.029***	-0.029***	-0.023**	-0.018*
	(0.007)	(0.005)	(0.005)	(0.007)	(0.008)
log(Startup Age)	-0.012	-0.002	-0.003	-0.040	-0.066**
	(0.023)	(0.024)	(0.024)	(0.032)	(0.026)
log(K)	-0.145***	-0.156***	-0.155***	-0.101***	-0.094***
	(0.021)	(0.017)	(0.017)	(0.027)	(0.023)
Time to Next Round	0.015***	0.018***	0.018***	0.019***	0.021***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Observations	5473	5468	5468	5468	5415
Adjusted R^2	0.018	0.019	0.020	0.035	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 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)	-0.007	-0.009	-0.009	-0.006	-0.009
	(0.008)	(0.007)	(0.007)	(0.008)	(0.009)
log(Startup Age)	-0.180***	-0.194***	-0.195***	-0.206***	-0.211***
	(0.018)	(0.020)	(0.022)	(0.021)	(0.021)
log(K)	-0.143***	-0.137***	-0.135***	-0.109***	-0.111***
	(0.008)	(0.007)	(0.010)	(0.014)	(0.017)
Time to Next Round	-0.003	-0.003	-0.002	-0.002	-0.002
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Observations	5473	5468	5468	5468	5415
Adjusted R^2	0.084	0.112	0.118	0.127	0.136
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

REFERENCES

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

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, 2022, 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.

Brueggeman, William B., and Jeffrey D. Fisher, 2019, *Real Estate Finance and Investments, 16th ed.,* McGraw-Hill Education.

Caliendo, Marco, and Alexander S. Kritikos, 2008, Is entrepreneurial success predictable? An ex-ante analysis of the character-based approach, *Kyklos* 61, 189–214.

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, 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.

Cole, Rebel A., Kenneth R. Ferris, and Arie Melnik, 2016, The direct cost of advice in M&A transactions in the financial sector, Working paper, Florida Atlantic University.

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.

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.

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

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

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.

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

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.

Lajoux, Alexandra R., and Charles M. Elson, 2010, *The Art of M&A Due Diligence, 2nd Edition,* McGraw-Hill Professional, New York.

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.

Ostrowski, Jeff, and Michele Petry, 2023, Is waiving a home inspection ever a good idea?, Bankrate.com.

NVCA, PitchBook, 2024, Venture capital monitor Q2 2024.

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

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

Testoni, Marco, Mariko Sakakibara, and M. Keith Chen, 2022, Face-to-face interactions and the returns to acquisitions: Evidence from smartphone geolocational data, *Strategic Management Journal* 43, 2669–2702.

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.

Appendix

Table 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 current round invested capital.
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 and all existing investors from previous rounds.

Abnormal Deal Volume	The number of deals for 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.
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.

Proofs

Proof of Prediction 1: Coming soon.

Proof of Prediction 2: Coming soon.

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

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

Combining the previous two relations,

$$\begin{aligned} \log(a) - \log(\hat{a}) &\sim N(\mu_1 - (\mu_1 + 1/(2\nu_1)), 1/\nu_1) && \text{(A.2)} \\ \log(a/\hat{a}) &\sim N(-1/(2\nu_1)), 1/\nu_1), && \text{(A.3)} \end{aligned}$$

and

$$\begin{aligned} E[MPK] &= E[a/\hat{a}] && \text{(A.4)} \\ &= \exp(-1/(2\nu_1) + 1/(2\nu_1)) = 1, && \text{(A.5)} \end{aligned}$$

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 $\Delta\Pi$:

$$\begin{aligned} \Delta\Pi &= (aK^{*\theta} - K^*) - (aK_i^{**\theta} - K_i^{**}) && \text{(A.6)} \\ &= a(K^{*\theta} - K_i^{**\theta}) - (K^* - K_i^{**}). && \text{(A.7)} \end{aligned}$$

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,

$$\Delta\Pi = aK^{*\theta}(1 - MPK^{\frac{\theta}{1-\theta}}) - K^*(1 - MPK^{\frac{1}{1-\theta}}). \tag{A.8}$$

Dividing both sides by K^*,

$$\frac{\Delta\Pi}{K^*} = aK^{*(\theta-1)}(1 - MPK^{\frac{\theta}{1-\theta}}) - (1 - MPK^{\frac{1}{1-\theta}}). \tag{A.9}$$

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

$$\begin{aligned} \frac{\Delta\Pi}{K^*} &= a(\theta\hat{a})^{-1}(1 - MPK^{\frac{\theta}{1-\theta}}) - (1 - MPK^{\frac{1}{1-\theta}}) && \text{(A.10)} \\ &= \frac{1}{\theta}MPK(1 - MPK^{\frac{\theta}{1-\theta}}) - (1 - MPK^{\frac{1}{1-\theta}}) && \text{(A.11)} \\ &= \frac{1}{\theta}(MPK - MPK^{\frac{1}{1-\theta}}) - (1 - MPK^{\frac{1}{1-\theta}}). && \text{(A.12)} \end{aligned}$$

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

$$\frac{d}{d\,(MPK)}\frac{\Delta\Pi}{K^*} = \frac{1}{\theta}\left(1 - MPK^{1/(1-\theta)-1}\right). \tag{A.13}$$

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

$$\frac{d^2}{d\,(MPK)^2}\frac{\Delta\Pi}{K^*} = -\frac{1}{1-\theta}MPK^{1/(1-\theta)-2}, \tag{A.14}$$

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 $\log(MPK^{\frac{1}{1-\theta}}) \sim N(\frac{\mu}{1-\theta}, \frac{\sigma^2}{(1-\theta)^2})$ and 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:

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

Combining terms produces equation (18).

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

$$\log(\tilde{V}) = \theta\log(K) + \log(a) + \log(\tilde{\epsilon}). \tag{A.15}$$

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:

$$\log(\tilde{V}) = \beta_0 + \beta_1 \log(K) + \alpha + \alpha_t + \delta, \tag{A.16}$$

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.15), we estimate θ as the coefficient β_1 in the OLS regression (A.16). 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.16) by OLS.

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

$$MPK = \theta \frac{V}{K} = \theta \frac{\tilde{V}}{K} \cdot \frac{1}{\tilde{\epsilon}}, \tag{A.17}$$

so

$$\log(MPK) = \log(\theta) + \log\left(\tilde{V}/K\right) - \log(\tilde{\epsilon}). \tag{A.18}$$

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.18) and taking variances of both sides yields

$$Var(\log(MPK)) = Var(\log(\tilde{V}/K)) - Var(\log(\tilde{\epsilon})). \tag{A.19}$$

We use the variance of the estimated residuals δ from regression (A.16) 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.19) produces an estimate of $Var(\log(MPK)) = \sigma^2$ equal to 0.54.

Online appendix

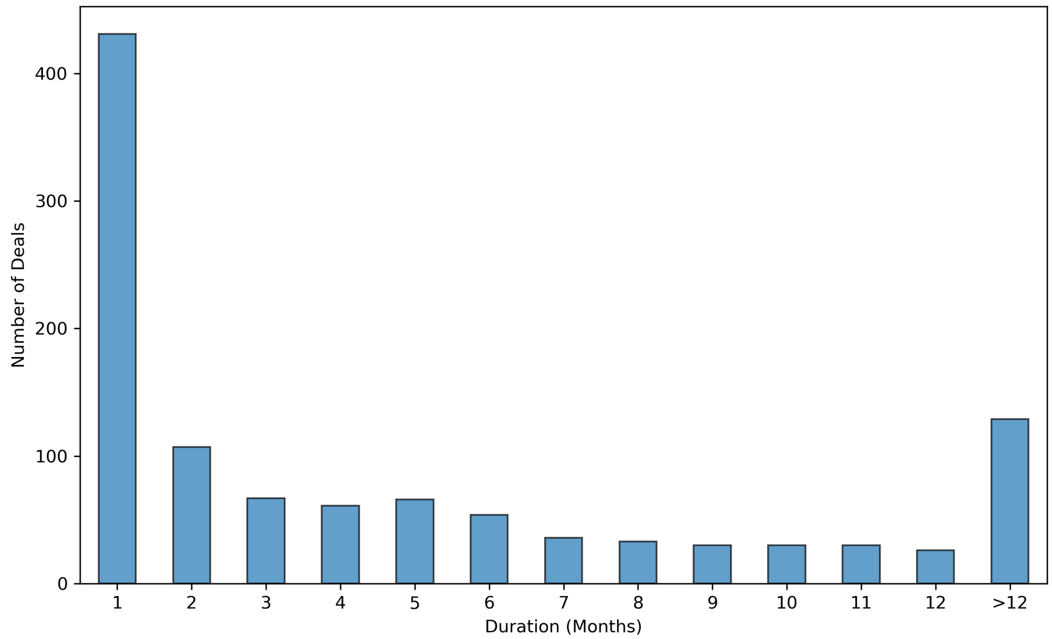

Figure B.1. Distribution of duration from pitch to investment. This figure shows the distribution of the duration from the pitch date to the investment date for deals with at least one captured pitch session within 18 months before the investment date. For deals with multiple pitch sessions, the last 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 samples have 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 of within 3 months.

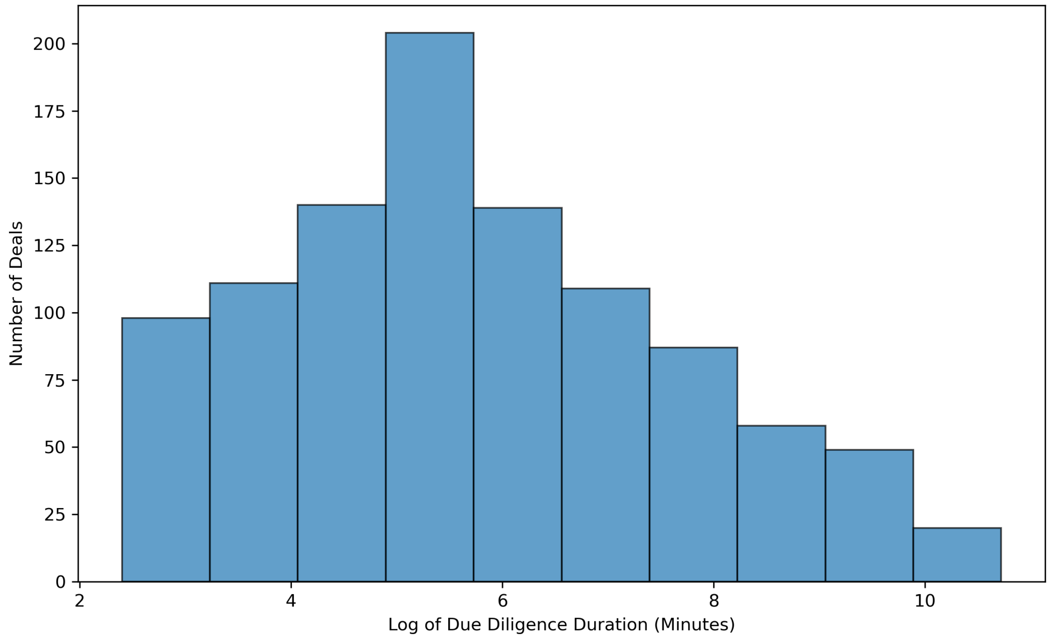

Figure B.2. Distribution of Due Diligence Duration. This figure plots the distribution of the log due diligence duration for deals with positive due diligence duration. Log transformation means $f(x) = \ln(1 + x)$ to avoid taking the logarithm of zero. All samples are between the lead investor and the startup, considering only new investors to the focal round. The 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 mitigate false positives, if a single meeting lasts shorter than 10 minutes or longer than 5 hours, it is considered a measurement error and deleted. When multiple employees are captured, the one with the longest duration is used. The mean due diligence duration corresponds to 5.4 hours, while the 25th, 50th (median), and 75th percentiles correspond to 1.3, 4.5, and 19.3 hours, respectively.

Table B.1
Robustness checks for Table 5

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. Asterisks *, **, 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**