NBER WORKING PAPER SERIES

COMMUNICATING PERFORMANCE: TEXTUAL AND FINANCIAL DISCLOSURES IN PRIVATE EQUITY

Josh Lerner
Fiona Paine
Antoinette Schoar

Working Paper 35860
http://www.nber.org/papers/w35860

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2026

We thank Tom Keck, Lisa Larsson, and Qi Liu of the StepStone Group for their support of the project. Financial support was provided by the Harvard Business School's Division of Research and Faculty Development. Josh Lerner has received compensation for advising institutional investors in private equity funds, private equity groups, and governments designing policies relevant to private equity. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w35860

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Josh Lerner, Fiona Paine, and Antoinette Schoar. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Communicating Performance: Textual and Financial Disclosures in Private Equity
Josh Lerner, Fiona Paine, and Antoinette Schoar
NBER Working Paper No. 35860
October 2026
JEL No. G24

ABSTRACT

Private capital fund managers appear to possess significant information about the relative quality of their portfolio firms even within the first three years after the investment. But the way this information is disclosed to investors changes dynamically: Textual discussions of portfolio companies have substantial information about the firms' ultimate performance in the early stages of investment, while interim valuations are less important. Over time, this pattern reverses, with interim valuations becoming more important in explaining future returns, especially after portfolio companies have a financing event. The results vary not only with the tone but also with the content of fund manager disclosures. In an opaque market with substantial information asymmetries, textual discussions seem to serve as a signal about general partner quality.

Josh Lerner
Harvard University
Harvard Business School
and NBER
jlerner@hbs.edu

Antoinette Schoar
Massachusetts Institute of Technology
MIT Sloan School of Management
and NBER
aschoar@mit.edu

Fiona Paine
Carnegie Mellon University
paine.f@gmail.com

1 Introduction

A large literature on delegated asset management has highlighted the central challenges of how managers report information about investment choices, fund performance, and the underlying risks to their investors. Rather than simply maximizing the informativeness of communication, managers might have incentives to disclose information strategically: for instance, limited disclosures may reduce competition from other fund managers or facilitate the raising of additional funds from investors.

These challenges are particularly acute in areas of asset management, such as venture capital and private equity, where disclosure is largely voluntary (at least at the level of individual companies in the portfolio). The non-compulsory nature of the disclosures about the firms in their portfolios gives general partners (GPs) significant discretion over the precision of the information that they disclose to their limited partners, as well as the timeliness and format in which information is made available. Moreover, the opacity of the underlying assets, the lack of independent analyst coverage, and illiquid secondary markets amplify the information gap between general partners and their limited partners (see Lerner et al. (2007)).

But it is ex ante not obvious *how* GPs should use this discretion. The literature on delegated asset management has identified several factors that can shape their incentives to report truthfully. On the one hand, general partners might face short-term incentives to overstate expected performance or underreport particularly poor-performing investments. Disclosing adverse information early might make it difficult to raise a new fund or even deter other entrepreneurs from accepting capital from them. Even if general partners are simply uncertain about the expected performance of firms in their portfolios, they might have incentives not to disclose much information early on or to strategically manipulate or obscure information—i.e., to "signal jam," as in Fudenberg and Tirole (1986) and Stein (1989)—to increase the chance that investors form a positive posterior of their underlying ability.

On the other hand, asset managers who systematically misreport the expected performance or risks in their funds might be punished in the medium run. For instance, they may be unable to raise a new fund or otherwise incur reputational damage. In an industry that is subject to significant underlying investment risks, and where even good managers make many investments that do not work out, GPs might have heightened incentives to signal their quality to their investors by providing detailed and accurate firm-level assessments as early as possible in the life of an investment (Leland and Pyle, 1977).

In this paper, we document whether and when private equity managers possess information about their portfolio companies, as well as the manner in which they disclose this information. The primary channel through which general partners communicate with their investors is through reports (or accompanying presentations) about their portfolio companies. These reports contain two broad types of information: financial disclosures, most notably interim valuations of each investment, and textual discussions of the portfolio companies' progress. Because private equity and venture capital investments lack observable market prices, interim valuations are largely GP-reported.[1] The textual disclosures capture more nuanced information about the risks, challenges, and opportunities of an investment.

We document a striking empirical pattern: General partners seem to have significant information about the relative quality of their portfolio firms, even very early in the life of an investment. Disclosures of interim financial performance are predictive of the investment's ultimate performance over the life of the fund. But interestingly, within the first three years of an investment, the sentiment expressed in GPs' textual discussions of portfolio companies contains substantial additional information about an investment's ultimate performance, while firms' interim valuations are somewhat less revealing. However, over time this pattern reverses and interim valuations become more important in explaining firms' ultimate

[1]While many private capital groups will have portfolio companies' valuations reviewed on an annual basis by auditors, there appears to be substantial leeway granted these groups when setting valuations. For one articulation of such concerns by Jay Clayton, former US Attorney for the Southern District of New York and US Securities and Exchange Commission chair, see https://www.fa-mag.com/news/private-credit-s-sketchy-marks-get-warning-shot-from-wall-street-s-top-cop-84983.html?section=43.

performance, while textual disclosures become unimportant.

Why might GPs choose to convey information through text rather than through returns, particularly when undervaluation could obscure evidence of fund performance and reduce positive attention from limited partners? Several features of private capital contracting make conservative interim valuations a rational—and in some cases, a preferred—strategy. First, the benefits to GPs of reporting higher valuations early in the fund's life are often modest. Management fees typically scale with committed, rather than reported, capital and carry is not realized until a successful exit. Many groups do not return to the market to raise funds until several years after the previous fundraising.[2] Second, the costs of overstated valuations can be large. Many endowments and pension funds compute distribution formulas based on net asset value; these distributions are difficult or impossible to reverse, even if later information reveals that interim valuations were inflated. In addition, GPs might worry that updating portfolio firm valuations early on, often well in advance of the company having an initial public offering or other liquidity event, might leak information to the outside world. The portfolio firm might consequently experience additional competition as other investors finance its competitors. In addition, the fund's bargaining power with the portfolio company might erode in the case of a refinancing.

Through this lens, GP narratives about portfolio firms offer a natural channel for conveying forward-looking assessments without the financial and contractual consequences associated with adjusting valuations. Consistent with this view, we find that sentiment predicts ultimate returns. This effect is primarily seen in the first few years after investment, precisely when valuation uncertainty is highest and when the costs of misvaluation are most asymmetric. As investments mature and uncertainty resolves, interim valuations increasingly incorporate the information reflected in the text, and the incremental predictive power of sentiment declines accordingly.

Our analysis uses data from StepStone Group, a large global private markets investment

[2]These periods may be when pressures to distort valuations are greatest, as Brown et al. (2019) document.

firm. We obtain end-of-year reports from general partners covering over 3,300 portfolio companies across 536 funds between 2015 and 2020. We chose this time period to ensure that we can observe the realized outcomes of the portfolio companies in the funds.

From these reports, we extract not only the interim valuations but also the text that GPs write about each portfolio company. We construct a measure of GP sentiment using a finance-specific dictionary (Loughran and McDonald, 2011) adapted for the private equity context, and BERTopic (Grootendorst, 2022) to classify the textual discussions into topic categories. This approach allows us to study both the tone and the content of GP communications at the individual portfolio company level, a level of granularity that has not previously been available.

Our sentiment index measures the normalized difference between positive and negative words in the GP's discussion of a specific portfolio company, standardized relative to the distribution across all company-years. We further construct measures of positive and negative sentiment separately, and use BERTopic—augmented with finance-specific embeddings—to classify sentiment-bearing sentences into topical categories such as Operations, Capital Structure, Supply Chain and Customer, Cash Flow, Macroeconomy, Leadership, and Business Model.

From these reports, we also extract free text and financial tables, from which we reconstruct standardized measures of firm-level revenues, cash flows, and valuations. These report-based data are then merged with external sources, including a proprietary fund-of-funds transaction database from StepStone containing detailed information on deal characteristics and exit outcomes, Preqin, and Capital IQ. We typically obtain an average of slightly more than two annual reports per portfolio company, with a higher density of reporting in the early years.[3] Individual write-ups contain roughly fourteen topic-tagged sentences, spanning three to four distinct thematic areas.

[3]Venture capital and private equity groups will typically report quarterly to their LPs. The year-end reports on which we focused are sometimes longer and are typically those that accompany the audited financial statements (though some funds are audited in a different quarter). In future work, we hope to analyze the full series of quarterly reports.

We first show that GP sentiment is indeed predictive of the ultimate performance of portfolio companies, but only in the early years of an investment. When we regress the realized (annualized) ultimate return of a portfolio company on the interim valuations and sentiment scores in the first three years after the initial financing, we find that a one standard deviation increase in sentiment is associated with an increase of three to four percentage points in final returns (annualized). This effect is meaningful relative to the sample mean of seven percent.

As the investment matures, however, sentiment's predictive power declines sharply and becomes indistinguishable from zero by year five and beyond. The coefficient on interim return moves in the opposite direction, rising sharply from the early years to the later periods. This pattern suggests that GPs possess private information from the outset that is initially reflected in their language but only gradually incorporated into reported valuations.

We also show that the presence of both optimistic and pessimistic language in the reports conveys information. There is little evidence of asymmetry in how GPs report information for portfolio firms that are expected to do well compared with those that are at the lower end of the performance spectrum.

Our findings are robust to controlling for potential confounding factors. One concern is that the declining predictive power of sentiment may simply be a consequence of the changing sample: companies that we observe in year five and beyond may be systematically different from the others. We restrict the sample to portfolio companies with multiple observations over time, and show that, here too, the patterns are consistent with information moving from qualitative disclosures to quantitative valuations as the investment matures and uncertainty about its prospects is resolved.

Nor is the predictive power of sentiment driven by stale valuations. Controlling for whether the valuation has been recently updated leaves the sentiment coefficient essentially unchanged. Winsorizing extreme returns confirms that a few outliers are not driving the results. Moreover, an interaction specification using the full sample yields consistent estimates

without requiring any sample splits.

We seek to shed more light on whether conservatism in writing up interim valuations explains the change in the predictive power of GPs' textual reports versus their interim valuations. Notably, we observe that financing rounds—which provide a market-validated price from an outside investor—reduce the predictive power of sentiment. When we break out our regressions of final returns on interim returns and the sentiment score, we see that the coefficient on sentiment scores becomes significantly smaller and insignificant after a portfolio firm has a financing event. These results suggest that GPs' textual communication is most informative when no independent valuation benchmark exists.

We also explore cross-sectional heterogeneity across fund types, seeking to understand differences in the information content of sentiment scores. We find suggestive evidence that sentiment scores are more predictive for larger funds (measured relative to fundraising in recent years), potentially reflecting their ability to provide more detailed information to their investors. We find no systematic differences by GP age or industry R&D intensity.

Finally, we examine not just the tone but also the content of GP discussions. We find that the choice of topic itself predicts performance. Early discussions focused on cash flow and business model are associated with better outcomes, while a focus on capital structure predicts worse performance. When we examine the predictive power of sentiment separately for each topic, the strongest predictive power comes from discussions of cash flow, macroeconomic conditions, and business model, while sentiment about leadership carries little information. The predictive power of macroeconomic sentiment is notable given that the regressions already control for interim return, suggesting that GPs may convey private assessments of how broader conditions affect their specific investments.

Taken together, the evidence supports a dynamic model of GP information transmission. Early in an investment's life, textual sentiment is a key vehicle for communicating private assessments, while interim valuations intentionally remain conservative (and less informative) due to institutional constraints. Over time, as uncertainty declines, valuations increasingly

reflect this information, reducing the incremental information from sentiment. These results suggest that GPs have incentives to signal their quality, i.e., the ability to assess portfolio company prospects, early in the investment cycle. However, as discussed above, they seem to be constrained in the manner in which they do this.

Our findings contribute to the literature on private-market reporting, valuation conservatism, and the informational content of qualitative disclosures. They offer new insights into how investors interpret GP communications in an environment of limited transparency.

In particular, a large body of research in private equity has examined the accuracy and informativeness of interim valuations, with particular attention to whether agency costs and career concerns lead GPs to manipulate reported values in order to raise follow-on funds (Jenkinson et al., 2013; Barber and Yasuda, 2017; Brown et al., 2019). A related literature has studied the relationship between interim valuations and subsequent realized cash flows (Jenkinson et al., 2020; Ercan et al., 2025), as well as the dynamics of markdowns and reputation (Hüther, 2023). Using unique textual and financial data from pre-investment deal plans, Biesinger et al. (2020) document the dimensions on which GPs select portfolio companies and how they plan to add value to the investments. There is also a smaller strand of literature suggesting that the illiquidity and infrequent repricing of private equity valuations may themselves serve a purpose: limited partners may prefer smooth reported returns, creating what has been called "volatility laundering" (Jackson et al., 2023). Yet across this body of work, the focus has been almost exclusively on the quantitative side of GP disclosures. The textual content of these reports has received little attention, mainly due to a lack of available data.

In public markets, however, there has been an increased awareness of the importance of textual communication as a complement to financial disclosures. Textual disclosures by managers of publicly traded corporations—with a focus on the sentiment, readability, and verifiability of the text—have been extensively scrutinized (see, for an overview, Loughran and McDonald, 2016). A growing body of work has demonstrated that the sentiment of

corporate communications and analyst reports contains information about future returns beyond what is captured in financial statements alone (Jiang et al., 2019). More recently, the role of narratives—which place corporate events in a broader context—in both corporate disclosures and analyst discussions has attracted increasing attention (Bybee et al., 2023; Flynn and Sastry, 2024; Sarkar, 2025).

A contemporaneous paper that is most closely related to our paper is Tamayo et al. (2026), which analyzes about 1,500 portfolio firms from Unigestion, a predominantly European private equity fund of funds. The authors also use FinBERT sentiment and find that a more positive report tone predicts higher subsequent returns. The two papers, however, complement each other in two main respects. First, the disclosure dynamics reported in Tamayo et al. (2026) are very different from ours. The authors find that in their sample, textual disclosures remain informative throughout the life cycle of the firm. We believe that the difference in the results might be driven by the fact that European private equity markets are less competitive than US markets (which are the focus of our study) and thus incorporate less information about a firm's ultimate quality. In addition, the fund and firm selection differs across the two samples. Unigestion is a significantly smaller fund of funds than StepStone and thus might draw from a different underlying firm distribution. Finally, we also decompose the textual content into the underlying topics that are being discussed by the GPs.

The remainder of the paper is organized as follows. Section 2 describes the data and the construction of our sentiment and topic measures. Section 3 presents the empirical analysis, including the baseline results, robustness checks, cross-sectional heterogeneity, and topic-level analysis. Section 4 concludes.

2 Data Construction

Our primary data source is the StepStone Group, a large global private markets investment firm responsible for approximately $811 billion of total capital, including $220 billion of assets under management.[4] StepStone collects detailed information on private equity funds, including those it invests in, those its limited partners (LPs) invest in, and funds that share information while soliciting new investments. We employ for this analysis the end-of-year (Q4) reports that StepStone receives from GPs on the funds in its portfolio. Our sample includes funds with vintage years between 2009 and 2018. This window was chosen to allow us to have several subsequent years to observe the performance of the investments that GPs make. These reports typically include not only fund-level performance but also information on individual portfolio companies. We focus only on US funds to reduce heterogeneity across regions, and, as a result, all text was in English. We supplement the report information, as noted above, with information from StepStone databases, as well as Preqin and Capital IQ.

From each end-of-year report, we extract three types of portfolio company-specific information: financial metrics such as revenue, profit, and/or cash flow; the reported valuation of the portfolio company; and—the major source of our contribution—the text that the GP wrote about each portfolio company.

The first step in constructing the dataset is to map each page of a GP report to the portfolio company discussed, discarding pages without portfolio company-specific information. For half the reports, we hand-label each page with the relevant portfolio company. For the remaining reports, we use an LLM-based BERT entity recognition model to identify the portfolio company discussed on each page.[5]

We test the accuracy of the automated approach using the hand-labeled dataset. On a sample of eight hand-labeled reports (with an average of 83 pages per report), the automated

[4] See https://www.stepstonegroup.com. Figures as of December 2025.

[5] Throughout, We processed the data in StepStone's cloud environment using locally hosted large language models from the Amazon Nova family. No data were sent out to external services, such as those operated by Anthropic or OpenAI.

approach identifies 104 out of 116 portfolio companies (89.7%). Of the pages automatically matched to companies, 91.7% are correct according to the hand-labeled dataset.

We use Amazon Textract to extract the text and tables from each labeled page of the GP reports, which are in PDF form. From each table, we extract the values and names of the variables reported, the currency (e.g., USD, CAD), units (e.g., thousands, millions), and for portfolio company cash flows, the reporting period (e.g., three-month, six-month, twelve-month). These individual tables are then combined into a single coherent dataset and merged with portfolio company and GP-level data from Preqin and Capital IQ, as well as data on portfolio company attributes and outcomes from the StepStone database.

Next, we use the portfolio company-specific text to create a measure of GP sentiment for each report about each portfolio company discussed. We use a dictionary-based approach, drawing on the finance-specific corpus of positive and negative words from Loughran and McDonald (2011). It is important to use a finance-specific corpus because general-purpose terms (e.g., "cancer") and financial terms (e.g., "debt") can be mislabeled as having sentiment when they are actually neutral in a finance context. Since our sample consists of venture capital and buyout deals, we modify the Loughran and McDonald (2011) dictionary by dropping terms that are not negative in a private equity/venture capital context. For example, we disregard the word "disrupt" since it is labeled as negative in Loughran and McDonald (2011) but can be positive in a venture capital setting. We also supplement the dictionary with private equity-specific terms drawn from the glossary of Lerner et al. (2023), which we hand-label as positive, negative, or neutral.

After labeling each word as positive, negative, or neutral, we calculate an overall measure of sentiment using the following equation:

$$\text{Net Sentiment} = \frac{\#\text{ Positive Words} - \#\text{ Negative Words}}{\#\text{ Total Words}} \tag{1}$$

We z-score normalize the sentiment measure across all portfolio company-years so that

the distribution has mean zero and standard deviation of one. A one-unit increase in the normalized sentiment thus corresponds to a one standard deviation increase in the net sentiment measure.

Since net sentiment captures the balance between positive and negative words, we also compute separate measures of positive and negative sentiment:

$$\text{Positive Sentiment} = \frac{\text{\# Positive Words}}{\text{\# Total Words}} \tag{2}$$

$$\text{Negative Sentiment} = \frac{\text{\# Negative Words}}{\text{\# Total Words}} \tag{3}$$

Finally, we extract not just overall sentiment but also the topics discussed in contextual snippets from GP reports. For expositional simplicity, we refer to these snippets as "sentences." We focus on sentences that contain sentiment-bearing language. For example, a sentence such as "The company operates in the healthcare sector" would not be included, while "The company experienced a significant decline in revenue" would be. To identify these sentences, we extract the 50 characters before and after each word labeled as positive or negative.

We then use the BERTopic model (Grootendorst, 2022) to group the sentiment-bearing sentences into topic categories. BERTopic generates document embeddings using pre-trained transformer-based language models, clusters these embeddings, and extracts topic representations using a class-based TF-IDF procedure. A key advantage of this approach is that it does not require pre-defining topic categories. To ensure that the resulting groupings reflect financial rather than industry-specific themes (e.g., a cluster of words about biotechnology), we use a finance-specific embedding model within BERTopic. Based on the resulting clusters, we assign seven topic categories: Operations, Capital Structure, Supply Chain and Customer, Cash Flow, Macroeconomy, Leadership, and Business Model.

3 Empirical Analysis

3.1 Summary Statistics

The dataset is at the portfolio company-by-annual report level. The main explanatory variables are derived from the periodic discussions of portfolio companies in the annual reports of the venture capital or private equity groups that backed them. As discussed above, we do not have annual reports for every year that a fund was active. As Table 1 reports, the median portfolio company is discussed in two annual reports, with a mean of 2.2 annual reports. The first report typically arrives one year after the initial investment, and the last report arrives around three years after the initial investment. Coverage declines in later years, either because the annual report was not available in our dataset or because the portfolio company was no longer discussed in the report.

Table 1 also reports the interim return (annualized from the time of the investment to the observation) and the final return (annualized from the time of the investment to the exit). The skewed nature of returns is evident: the mean interim return of 0.20 substantially exceed the median of 0.11, reflecting a long right tail of high-performing investments. Overall return levels, even after adjustments for fees and carry, are above typical industry benchmarks, suggesting either outperformance among StepStone's portfolio companies or a bias on the part of GPs toward reporting on successful firms (or both). The corresponding sentiment scores, reported in Table 2, are centered near zero (mean of -0.01, standard deviation of 0.94), as expected from the z-score normalization.

The distributions of sentiment and interim returns are shown in more detail in Figures 1 through 3. Sentiment follows an approximately bell-shaped distribution for both early and late observations. The within-company change in sentiment (the difference between last and first observation) is tightly concentrated around zero, indicating that GPs' tone toward a given portfolio company remains relatively stable over time.

Interim returns (not annualized), by contrast, shift over the life of the investment. As

Figure 2 shows, early observations are clustered near zero, reflecting valuations that have not yet moved far from cost, while later observations are more dispersed as outcomes become clearer. When annualized in Figure 3, the early and late distributions look much more similar, as the mechanical effect of compounding over longer horizons is removed.

Table 2 takes a deeper look at the text discussions of portfolio firms. The median write-up contains roughly 351 words, with about eleven sentences per write-up containing sentiment-bearing language that can be assigned to a topic. Each write-up typically covers three-and-a-half topics, with about half the space devoted to the most important one.

Table 3 provides an overview of the dataset coverage. The sample contains 7,271 observations spanning 3,313 portfolio companies across 536 funds, covering the period from 2015 to 2020. The sample is predominantly of buyout transactions, with venture capital observations accounting for only 14% of the total.[6] In roughly 20% of cases where an annual report is available and the investment is still active, the portfolio company is not discussed. As we show in the Appendix (Tables A5–A9), these omissions are not systematically related to portfolio company performance, though they are negatively related to sentiment.

Table 4 describes the characteristics of the funds in our sample. The median fund is the sixth for the private capital group. The median group was established in 1996. The median fund size is $1.2 billion.[7] Median net IRR is 17.2 percent and median DPI is 130%, indicating that these are mature, well-performing funds managed by experienced general partners.

3.2 Performance and Sentiment

The key results of the paper are presented in Table 5. Our baseline regression specification, which forms the foundation for much of the subsequent analysis, is as follows:

$$\text{Final Return}_i = \alpha + \beta_1 \cdot \text{Sentiment}_{i,t} + \beta_2 \cdot \text{Interim Return}_{i,t} + \gamma_f + \epsilon_{i,t} \tag{4}$$

[6] Our main results are generally robust to treating venture capital observations distinctly (see Appendix Table A1) or excluding them entirely (Tables A2 through A4 in the Appendix).

[7] Median natural log of fund size is 7.09. Thus, because fund size is measured in millions, $\exp(7.09) \approx 1.2$ billion.

where the dependent variable is the annualized final return of portfolio company i from the time of the investment to the exit, $\text{Sentiment}_{i,t}$ is the z-scored net sentiment measure as of year t, $\text{Interim Return}_{i,t}$ is the annualized interim return as of the report date, and γ_f denotes fund fixed effects (included in some specifications). We split the sample into observations from annual reports in three time intervals: years zero through two after the initial closing, years three and four, and year five and beyond. We estimate each specification with and without fund fixed effects.[8]

The results show that both interim return and sentiment have predictive power for ultimate returns, but their relative importance varies over the life of the investment. In the early years (Columns 1 and 4), sentiment is positive and statistically significant: a one standard deviation increase in sentiment is associated with an increase of four percentage points in final returns. The estimated effect is economically meaningful relative to the sample mean of 0.07. This result suggests that interim valuations do not fully capture the information held by general partners, with additional information conveyed through the language of their reports. As the investment matures, however, the sentiment coefficient declines from 0.04 in years zero through two to 0.02 in years three and four, and becomes indistinguishable from zero by year five and beyond. The coefficient on interim return moves in the opposite direction, rising from 0.16 in the early years to 0.62 in years three and four and 0.78 in year five and beyond. The patterns are broadly similar when fund fixed effects are included (Columns 4 through 6), with the coefficient on sentiment equal to 0.03 in the early years and interim return rising from 0.16 to 0.72.

In Table 6, we estimate the same specification as in Equation 4, but decompose net sentiment into its positive and negative components.[9] Odd-numbered columns use positive sentiment while even-numbered columns use negative sentiment. Both follow the same pat-

[8] Our results are robust to splitting the sample at the year level (Table A15), and to using returns that have not been annualized (Table A16).

[9] In Appendix Table A10, we also run an alternative specification, splitting the sample by whether sentiment is above or below median *within the group* in the year of the investment. Here the results are much smaller in magnitude.

tern as the net measure: in the early years, positive sentiment is associated with higher final returns (0.03) while negative sentiment is associated with lower final returns (−0.02), with both statistically significant. By years three and four, the coefficients shrink roughly by half (0.01 and −0.01, respectively). By year five and beyond, neither is distinguishable from zero. The coefficients on interim return are essentially unchanged from Table 5 throughout.

This analysis confirms that the net sentiment result is not driven by one side alone. Both the presence of optimistic language and the presence of pessimistic language convey information about ultimate outcomes in the early years of an investment.

3.3 Robustness Checks

We now examine several potential explanations for these patterns in a series of robustness checks.

A natural concern is that our findings may reflect stale valuations rather than private information. If general partners do not update their interim valuations frequently, the reported valuation may lag the GP's actual assessment of the investment. In that case, sentiment could appear to predict final returns simply because it reflects information that the GP already has but has not yet incorporated into the stated valuation.

There are several forces that push general partners to maintain current valuations. The US Generally Accepted Accounting Principles require private equity groups to use Accounting Standards Codification 820. This provision governs the determination of fair value of investments, instructing firms to value such stakes at the "price that would be received in an orderly transaction using market participant assumptions at the valuation date." Meanwhile, the US Securities and Exchange Commission has promulgated increasingly stringent rules stipulating the need for regular disclosures of valuations (though some have subsequently been repealed). Nonetheless, numerous practitioner accounts suggest that "stale prices," or dated valuations, remain a common feature of the private equity landscape.[10] The prevalence

[10] A recent example is the discussion by Cliff Asness in https://www.capitalallocators.com/podcast/

of stale pricing despite these regulatory pressures makes this concern particularly relevant for our setting.

To address this possibility, Table A11 adds an indicator variable for whether the valuation has been updated, defined as a change of more than one percent from the prior year. Columns (1) through (3) add this indicator to the baseline specification with fund fixed effects, while Columns (4) through (6) further interact the update indicator with interim return, to test whether the predictive power of interim returns differs depending on whether the valuation has been recently updated.

In years zero to two, valuation updates are themselves associated with higher final returns (0.08), suggesting that these events—often tied to supplemental financing rounds—carry positive information about the investment's trajectory. This effect fades in later periods, with the update coefficient dropping to 0.05 in years three and four and becoming essentially zero by year five and beyond. Crucially, the sentiment coefficient remains at 0.04 in the early years, essentially unchanged from the baseline estimates in Table 5, and the interim return coefficient follows a pattern similar to the baseline estimates, rising from 0.14 in the early years to 0.72 by year five and beyond. In Columns (4) through (6), the interaction between the update indicator and interim return is insignificant across all time periods, and sentiment remains significant at 0.04 in the early years. Overall, these results indicate that the predictive power of sentiment is not driven by stale valuations: even after controlling for whether the valuation has been recently updated, the language of the GP's report contains additional information about ultimate performance.

A second concern is that the declining predictive power of sentiment may reflect changing sample composition rather than a genuine shift in informativeness. The baseline analysis in Table 5 uses all available observations, but the set of companies observed in the early years need not be the same as those observed later. Some companies may exit early through an IPO or acquisition and never appear in the later time periods, while others may only

`simple-investing-is-hard/`. This issue has been explored extensively in the literature on risk-adjustment of private equity returns (for a review, see Korteweg 2019).

begin to be discussed after several years. If the companies that remain in the sample at year five and beyond are systematically different—for instance, if they are investments with more uncertain outcomes where sentiment was never particularly informative—then the declining coefficient on sentiment could be an artifact of which companies appear in each time period, rather than a change in sentiment's predictive power over time.

To address this concern, Table 7 restricts the sample to portfolio companies with at least two observations separated by at least two years. We impose this restriction on the data to ensure that there is a meaningful time gap between the first and the second report, which increases the likelihood that new information about the companies becomes available.

In Columns (1) through (3), we split this restricted sample by time period as before. The same pattern holds: sentiment is 0.04 and significant in the early years, but falls to 0.003 in years 3 and 4 and loses significance. Meanwhile the interim return coefficient grows in magnitude over time. This trend continues for the observations in years 5 and later. This analysis confirms that the baseline results are not driven by companies that appear only once in the sample.

Columns (4) and (5) then isolate the first and last observations for each company in this restricted sample and estimate separate regressions on each. Because the same 1,102 companies appear in both columns, any difference in coefficients cannot be attributed to changes in the sample composition. In Column (4), we only look at the sentiment score and interim returns based on the first GP report for a company. The coefficient on sentiment is 0.02 but is not statistically significant. The coefficient on the Interim Return is 0.14 and significant at the one percent level. When we repeat this regression in Column (5) for the last GP report about these same firms, however, the coefficient on sentiment drops to 0.006 but is not significantly different from the sentiment coefficient in (4) ($p = 0.37$). The interim return coefficient, by contrast, increases sharply from 0.14 to 0.84. This pattern suggests that interim valuations increasingly incorporate more information, reducing sentiment's incremental contribution relative to the stated valuation.

In Column (6), we examine more directly whether information initially conveyed through the textual description subsequently appears in the interim valuations. The dependent variable here is the last observed interim return of a company. The key explanatory variable is sentiment in the company's first observed report. The coefficient on first sentiment is 0.04 and statistically significant at the one percent level. Thus, a one-standard-deviation increase in sentiment in the first report predicts a four-percentage-point higher annualized interim return in the final report that we observe for that company. This result provides direct evidence that the information embedded in early textual disclosures is subsequently reflected in the GP's quantitative valuation of the investment.

In Column (7), we then regress final return on sentiment in the first report, sentiment in the last report, and the last interim return. The coefficient on the last interim return is 0.84 and highly statistically significant. The coefficients on first and last sentiment are -0.01 and 0.01, respectively, and neither is statistically significant. Taken together, Columns (6) and (7) sharpen the dynamic interpretation of the baseline findings. Early in the investment, textual disclosures contain information that is not yet fully incorporated in interim valuations. That information predicts subsequent valuation changes, but once the final interim valuation is observed, neither early nor contemporaneous sentiment provides additional explanatory power for realized returns. The evidence is therefore consistent with information moving from qualitative disclosures into quantitative valuations as the investment matures and uncertainty about its prospects is resolved.

As a further check on the sample composition concern, Table 8 estimates the baseline specification on the full sample of 7,271 observations, interacting both sentiment and interim return with time period indicators (with years zero through two as the omitted reference category). This approach avoids splitting the sample, allowing all observations to contribute to the estimation.[11] The base sentiment coefficient is 0.04, consistent with the split-sample results. The interaction terms show that sentiment's predictive power declines by -0.02 in

[11] As an additional check, we also estimate the baseline specification interacting with the time period group as a continuous variable in Appendix Table A12.

years three and four and by −0.04 in year five and beyond, both significant at the one percent level. Interim return moves in the opposite direction, with interaction terms of 0.46 and 0.62 for the later periods. These results are robust to the inclusion of fund fixed effects (Column 2), where the patterns are nearly identical. Column (3) includes portfolio company fixed effects, which absorb nearly all variation. The within R^2 drops to 0.004, as identification in this specification relies on within-company changes in sentiment over time, which as Figure 1 showed are quite limited.

A third concern is that the results may be driven by a few outliers. Private equity returns are highly skewed (Brown et al., 2020), with a long tail of positive outcomes and some complete losses. Therefore, it is natural to wonder whether these extreme observations may be driving the relationship between sentiment and final returns.

In response to this concern, Table A13 re-estimates the baseline specification for the early years, the period where sentiment is most predictive, under three winsorization schemes: capping the top five percent of final returns (Column 1), setting negative returns to zero (Column 2), and applying both outlier replacements (Column 3). The sentiment coefficient ranges from 0.02 to 0.03 across the three specifications, remaining positive and statistically significant in all cases. The interim return coefficient is also stable, ranging from 0.09 to 0.14. These results confirm that the baseline findings are not driven by extreme outcomes at either tail of the return distribution.

3.4 General Partners, Financing Events, and Industry Conditions

Having established that the baseline relationship between sentiment and final returns is robust to several potential concerns, we now explore cross-sectional heterogeneity in the information content of sentiment.

We begin by examining whether the predictive power of sentiment varies across general partners. Earlier work has documented behavioral differences between more and less established fund managers, on dimensions including initial valuation levels (Hsu, 2004), exit

timing (Gompers, 1996), and post-IPO performance (Krishnan et al., 2011). More established groups—proxied by larger funds or older organizations—might feel less pressure to engage in strategic behavior vis-à-vis their limited partners (Gompers, 1996). If so, these groups would have less incentive to communicate through indirect channels like sentiment, and their interim valuations would more fully reflect their private information, leaving less for sentiment to predict. Alternatively, if established groups do convey information through their language, their deeper understanding of investment dynamics may make their sentiment more predictive of future performance. We term these the strategic behavior and information stories, respectively.

Table 9 examines whether the predictive power of sentiment and interim return varies with GP characteristics. In Columns (1) through (3), we interact sentiment and interim return with the log of adjusted GP fundraising, calculated as the log of the fund's size normalized by the prior five-year average of global GP fundraising. This measure captures the relative scale of the fund compared with its industry peers. In the early years, the interaction between sentiment and fund size is positive (0.008) and marginally significant, suggesting that sentiment is slightly more predictive for larger normalized funds. The interaction between interim return and fund size is only marginally significant in the middle period.

In Columns (4) through (6), we interact our measures with the number of years that the GP has been active, measured as of the end of our sample period. The interaction between interim return and GP age is negative (-0.002) and marginally significant in the early years, suggesting that older GPs' valuations are somewhat less predictive early on, possibly reflecting more conservative early valuations. The corresponding sentiment interactions are small and not significant in the early or middle periods, though marginally positive (0.001) in year five and beyond. Taken together, the results lean slightly toward the information story—that larger, more established groups may produce more informative assessments of their portfolio companies—though the evidence is not strong enough to draw firm conclusions.

We next look at whether specific valuation events reduce the informational content

of sentiment. A natural candidate is a follow-on financing round, which provides a market-validated price, typically from an outside investor, and represents one of the most informative valuation signals in private equity. If sentiment's predictive power stems from the gap between the GP's private assessment and the stated valuation, then a financing round—by forcing the valuation to update—should reduce this gap and diminish sentiment's incremental contribution.

Table 10 tests this suggestion in two ways. In Columns (1) through (3), we interact sentiment and interim return with an indicator for whether the observation occurs after the portfolio company has raised a follow-on round (Post Round Ind.). In Columns (4) through (6), we use an indicator for whether the observation falls in the same year as a funding round (Round Year).

The key results concern the interaction between financing events and sentiment. In years three and four, the Post Round interaction with sentiment is -0.04 and is significant at the five percent level (Column 2), while the Round Year interaction with sentiment is -0.07, again significant at the five percent level (Column 5). Both coefficients suggest that financing rounds partially substitute for the information conveyed through GP language. In the early years, neither interaction is significant: the base sentiment coefficient remains at 0.03 in both specifications, suggesting that even a market-validated round price does not fully resolve the uncertainty about the investment at that stage. By year five and beyond, sentiment has already lost its predictive power, leaving little room for financing events to have an impact.

Finally, we examine whether the information content of sentiment varies with the degree of uncertainty in the portfolio company's industry. In sectors with greater information asymmetry, general partners may be more reluctant to update valuations, due to the difficulty of determining fair value. In these cases, the gap between the GP's private assessment and the stated valuation may be wider. Sentiment may be more predictive in these sectors. To proxy for industry-level uncertainty, we use sector R&D intensity. Sectors with high

R&D spending tend to have greater uncertainty about future cash flows, as the outcomes of research and development activities are inherently difficult to predict. We calculate R&D intensity as the average ratio of R&D expense to revenue among public companies in the same two-digit GICS sector between 2005 and 2015, using data from Compustat.

Table A14 adds this proxy to the baseline specification, interacting both sentiment and interim returns with sector R&D intensity. Contrary to the hypothesis, sentiment is not more predictive in R&D-intensive sectors. The interaction between sentiment and R&D intensity is negative across all periods and not significant. Meanwhile, the interaction between interim return and R&D intensity is positive (2.1) and marginally significant in years three and four, suggesting that valuations are actually more informative in these sectors. The base coefficients on sentiment and interim return remain consistent with earlier tables (0.04 and 0.14 in the early years, respectively).

One possible explanation is that in R&D-intensive industries, valuation events—such as clinical trial results, patent grants, or new product launches—are discrete and highly informative. When valuations do update in these sectors, they carry strong signals about the investment's trajectory, making interim returns a more powerful predictor. At the same time, GP language about R&D-intensive companies may be less informative if it relies on more speculative or forward-looking assessments that are harder to map onto ultimate outcomes.

3.5 The Nature of Textual Discussion

As a final set of analyses, we look at the nature of the discussion by the general partners in more detail. We seek to understand whether the content of these discussions, which (as the reader will recall from Section 2) we divided into seven topics, is related to performance.

Table 11 presents an initial analysis, in which we add topic indicator variables to the baseline specification from Table 5. Each indicator equals one if the corresponding topic has the most sentences in that year's report, controlling for overall sentiment and interim return. This approach allows us to test whether the subject matter of the GP's discussion predicts

performance independently of tone.

In the early years, discussions focused on cash flow and business model (both with a coefficient of 0.05) are associated with significantly better ultimate performance. A focus on capital structure is associated with poorer outcomes. Early attention to business fundamentals may signal a well-positioned investment, while a focus on debt and financing may reflect structural challenges facing the company. Cash flow stands out as the one topic that is consistently associated with better performance across both early and late periods (0.05 and then 0.03), suggesting that GP discussions of cash generation are informative regardless of the investment's stage. By year five and beyond, topics that were negative early on—capital structure and supply chain and customer—become positively associated or not significantly associated with performance, possibly reflecting a shift in context as these discussions relate more to exit preparation than to business fundamentals. The coefficients on sentiment and interim return in this specification are essentially unchanged from Table 5, indicating that topic choice carries information beyond what is captured by tone or valuations alone.

While Table 11 shows that the choice of topic itself predicts performance, it does not tell us whether the tone of these discussions matters differently across topics. We now turn to this question. Table 12 focuses on early observations (years zero through two) where a particular topic was the main topic discussed. We further compute sentiment using only sentences about that topic. In Table 13, we take the sample of portfolio companies where a particular topic was the main topic discussed early and examine the full set of observations for the portfolio company. Sentiment is calculated using the complete text. We interact sentiment and interim return with an indicator for whether the observation falls in the early period.

The predictive power of sentiment documented earlier is not uniform across topics. In Table 12, the strongest sentiment coefficients in the early years are for macroeconomy (0.13), cash flow (0.12), and business model (0.10). Capital structure, operations, and leadership sentiment carry little or no predictive power.

Table 13 confirms this pattern using the full sample. The Sentiment × Early Observation interaction is significant for macroeconomy (0.05), operations (0.05), cash flow (0.03), and marginally significant for business model (0.03). In terms of interim returns, the pattern is more universal. The Interim Return × Early Observation interaction is negative and significant across all topics (ranging from −0.21 to −0.45), confirming that interim return becomes more predictive in later periods regardless of the subject matter being discussed.

These findings suggest that the information content of GP language is concentrated in specific areas. That cash flow sentiment is a strong predictor is consistent with GPs possessing private information that is most valuable when it pertains to company-specific fundamentals. The predictive power of macroeconomic sentiment is notable, given that the regressions already control for interim return. To the extent that publicly available macroeconomic information is reflected in valuations, the residual predictive power of macroeconomy sentiment suggests that GPs may be conveying private assessments of how broader conditions affect their specific portfolio company. Sentiment about leadership, by contrast, carries little predictive power.

To directly measure how topics co-move with sentiment, Table A17 studies how changes in the intensity of a topic's discussion correlate with sentiment. Increases in the discussion of the macroeconomy and capital structure correlate with a decrease in text sentiment. Increase in the intensity of discussion about operations and business model correlate with an increase in sentiment. This suggests that sentiment is driven by changes in the information communicated by the GP. Moreover, certain topics are not neutral: they are more likely to be either blamed (macroeconomy) or given credit (business model) when a GP is discussing a portfolio company. An important caveat is that the topics which drive sentiment may or may not actually contribute to the predictive power of sentiment. However, as was shown previously, macroeconomy and business model do contribute to the predictive power of sentiment.

4 Conclusion

In this paper, we use a novel dataset from the StepStone Group to explore communications from GPs to their investors. We show that private capital fund managers appear to possess substantial information about the relative quality of the firms in their portfolios. In the three years after the investment, their textual discussions of portfolio companies predict the firms' ultimate performance. Interim valuations also reveal some information, but are less important. Over time, this pattern reverses, with interim valuations having much more explanatory power for future returns. The results are robust to a variety of alternative explanations, and also vary with the content of fund manager discussions.

The analysis suggests the potential to glean additional information from these communications by GPs to their LPs. Given the opaqueness of private markets, one might anticipate that textual analysis could shed light on the ways in which GPs react to changes in market conditions or technological shocks, in a way that would complement existing studies analyzing investment behavior or survey responses. Novel language associated with such shocks (building on the approach of Hassan et al. (2019) and Kalyani et al. (2025), among others) might shed new light on private markets.

References

Barber, B. M. and Yasuda, A. (2017). Interim Fund Performance and Fundraising in Private Equity. *Journal of Financial Economics*, 124(1):172–194.

Biesinger, M., Bircan, C., and Ljungqvist, A. (2020). Value Creation in Private Markets. *CEPR Discussion Paper*, DP14676.

Brown, G. W., Gredil, O. R., and Kaplan, S. N. (2019). Do Private Equity Funds Manipulate Reported Returns? *Journal of Financial Economics*, 132(2):267–297.

Brown, G. W., Harris, R. S., Hu, W., Jenkinson, T., Kaplan, S. N., and Robinson, D. T. (2020). Private Equity Portfolio Companies: A First Look at Burgiss Holdings Data. Unpublished working paper.

Bybee, L., Kelly, B., and Su, Y. (2023). Narrative Asset Pricing: Interpretable Systematic Risk Factors from News Text. *Review of Financial Studies*, 36(12):4759–4787.

Ercan, E. Y., Kaplan, S. N., and Strebulaev, I. A. (2025). Interim Valuations, Predictability, and Outcomes in Private Equity. National Bureau of Economic Research Working Paper No. 33637.

Flynn, J. P. and Sastry, K. (2024). The Macroeconomics of Narratives. Unpublished working paper.

Fudenberg, D. and Tirole, J. (1986). A "Signal-Jamming" Theory of Predation. *RAND Journal of Economics*, 17(3):366–376.

Gompers, P. A. (1996). Grandstanding in the Venture Capital Industry. *Journal of Financial Economics*, 42(1):133–156.

Grootendorst, M. (2022). BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure. *arXiv preprint arXiv:2203.05794*.

Hassan, T. A., Hollander, S., van Lent, L., and Tahoun, A. (2019). Firm-Level Political Risk: Measurement and Effects. *Quarterly Journal of Economics*, 134(4):2135–2202.

Hsu, D. H. (2004). What Do Entrepreneurs Pay for Venture Capital Affiliation? *Journal of Finance*, 59(4):1805–1844.

Hüther, N. (2023). Do Private Equity Managers Raise Funds on (Sur)Real Returns? Evidence from Investment-Level Data. *Journal of Financial and Quantitative Analysis*, 58(7):2959–2992.

Jackson, M. O., Ling, D. C., and Naranjo, A. (2023). Volatility Laundering. Unpublished working paper.

Jenkinson, T., Landsman, W. R., Rountree, B. R., and Soonawalla, K. (2020). Private Equity Net Asset Values and Future Cash Flows. *Accounting Review*, 95(1):191–210.

Jenkinson, T., Sousa, M., and Stucke, R. (2013). How Fair are the Valuations of Private Equity Funds? Unpublished working paper.

Jiang, F., Lee, J., Martin, X., and Zhou, G. (2019). Manager Sentiment and Stock Returns. *Journal of Financial Economics*, 132(1):126–149.

Kalyani, A., Bloom, N., Carvalho, M., Hassan, T. A., Lerner, J., and Tahoun, A. (2025). The Diffusion of New Technologies. *Quarterly Journal of Economics*, 140(2):1299–1365.

Korteweg, A. (2019). Risk Adjustment in Private Equity Returns. *Annual Review of Financial Economics*, 11:131–152.

Krishnan, C., Ivanov, V. I., Masulis, R. W., and Singh, A. K. (2011). Venture Capital Reputation, Post-IPO Performance, and Corporate Governance. *Journal of Financial and Quantitative Analysis*, 46(5):1295–1333.

Leland, H. E. and Pyle, D. H. (1977). Informational Asymmetries, Financial Structure, and Financial Intermediation. *Journal of Finance*, 32(2):371–387.

Lerner, J., Leamon, A., and Hardymon, F. (2023). *Venture Capital, Private Equity, and the Financing of Entrepreneurship*. John Wiley & Sons, 2nd edition.

Lerner, J., Schoar, A., and Wongsunwai, W. (2007). Smart Institutions, Foolish Choices? The Limited Partner Performance Puzzle. *Journal of Finance*, 62(2):731–764.

Loughran, T. and McDonald, B. (2011). When is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance*, 66(1):35–65.

Loughran, T. and McDonald, B. (2016). Textual Analysis in Accounting and Finance: A

Survey. *Journal of Accounting Research*, 54(4):1187–1230.

Sarkar, S. K. (2025). Economic Representations. Unpublished working paper.

Stein, J. C. (1989). Efficient capital markets, inefficient firms: A model of myopic corporate behavior. *Quarterly Journal of Economics*, 104(4):655–669.

Tamayo, B. F., Braun, R., Lopez-de-Silanes, F., Phalippou, L., and Sigrist, N. (2026). Would I Lie to You? On Private Equity Intermediary Performance Reports. Unpublished working paper.

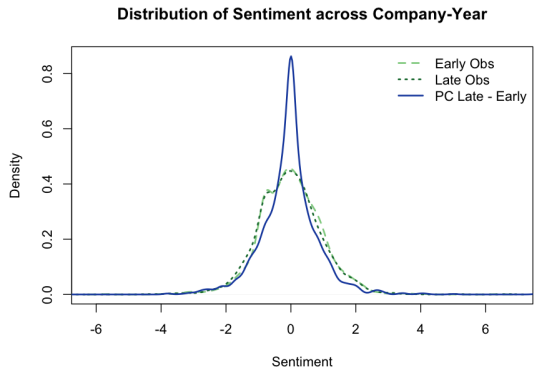


Figure 1: **Distribution of the sentiment variable.** The sample consists of cases where there are at least two observations for each portfolio company. The light green dashed line is the distribution for the first observation from each portfolio company. The dark green dotted line is the distribution for the last observation from each portfolio company. The dark blue solid line is the within-company difference between sentiment in the last and first observations. Sentiment is calculated as the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1.

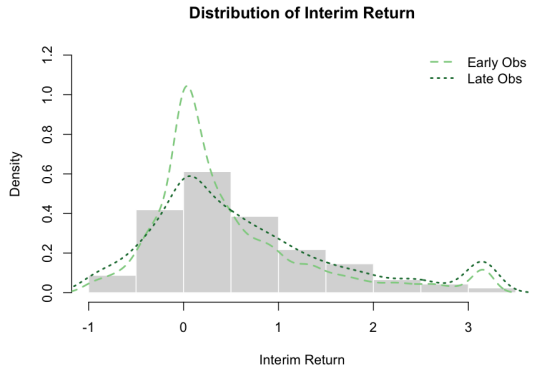


Figure 2: **The distribution of interim return.** The interim return is defined as the return from entry through observation time t. The light green dashed line is the distribution for the first observation from each portfolio company. The dark green dotted line is the distribution for the last observation from each portfolio company. The histogram bars show the distribution of the average interim return at the portfolio company level.

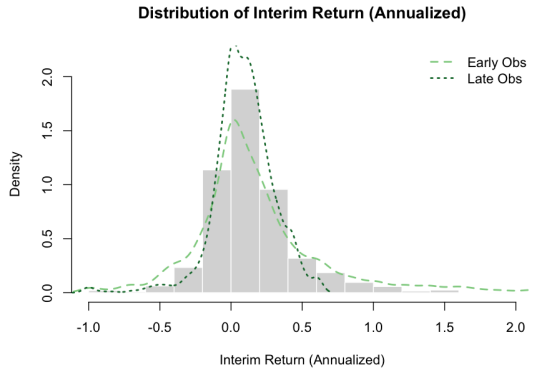


Figure 3: **The distribution of interim return (annualized).** In this figure, the interim return is the annualized return from entry through observation time t. The light green dashed line is the distribution for the first observation from each portfolio company. The dark green dotted line is the distribution for the last observation from each portfolio company. The histogram bars show the distribution of the average interim return (annualized) at the portfolio company level.

	p25	median	mean	p75	sd
Observations per Portfolio Company	1	2	2.19	3	1.45
Time of First GP Report	1	1	1.86	2	1.32
Time of Last GP Report	2	3	3.04	4	1.81
Interim Return	−0.001	0.11	0.20	0.30	0.47
Final Return	−0.01	0.10	0.07	0.25	0.35
log(Revenue)	3.72	4.89	4.89	6.22	2.93

Table 1: **Summary of financial reporting.** Observations per Portfolio Company is the number of observations we have for each portfolio company. Time of First GP Report is the number of years between the initial investment and the first observation of a portfolio company. Time of Last GP Report is the number of years between when the investor invests in a portfolio company and the last observation of that portfolio company. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company. log(Revenue) is the log of the reported portfolio company revenue, where revenue is in millions of USD.

	p25	median	mean	p75	sd
Sentiment	−0.61	−0.04	−0.01	0.56	0.94
Number of Positive Words	3	7	8.99	12	8.93
Number of Negative Words	1	2	3.50	5	4.13
Total Number of Words	214.88	351	416.70	546	298.58
Total Unique Topics	2	3	3.53	5	2.02
First GP Report Total Unique Topics	2	4	3.53	5	2.11
Last GP Report Total Unique Topics	2	4	3.52	5	2.13
Max Topic Share	0.36	0.46	0.52	0.62	0.22
Total Topic Sentences	5	11	14.31	19	13.29

Table 2: **Summary of textual reporting.** Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize this measure across all portfolio company years, so the distribution of sentiment scores has mean zero and a standard deviation of 1. Number of Positive Words is the number of positive words in a GP's report about a portfolio company. Number of Negative Words is the number of negative words in a GP's report about a portfolio company. Total Number of Words is the total number of words in a GP's report about a portfolio company. Total Unique Topics is the number of unique topics discussed in a given portfolio company report. First GP Report Total Unique Topics is the number of unique topics discussed in the first portfolio company report we have in our sample. Last GP Report Total Unique Topics is the number of unique topics discussed in the last portfolio company report we have in our sample. Max Topic Share is the largest share of sentences dedicated to a topic in a given portfolio company report. Total Topic Sentences is the total number of sentences assigned to any topic in a given portfolio company report.

	Value
Number of Observations in Sample	7,271
Number of Portfolio Companies in Sample	3,313
Number of Funds in Sample	536
First Year of Sample	2015
Last Year of Sample	2020
Years of Data in Sample	6
Ratio of Missing to Expected Obs.	0.20
Ratio of VC Observations	0.14

Table 3: **Summary of sample.** The table shows information about our dataset coverage. Ratio of Missing to Expected Obs. is the share of expected portfolio company observations for which a GP report is available but the company is not discussed. Ratio of VC Observations is the number of observations in the sample from a venture capital investor relative to the entire sample.

	p25	median	mean	p75	sd
Fund DPI (%)	89.70	130.07	141.13	180	84.88
Log(Fund Size)	6.26	7.09	7.18	8.26	1.49
Net IRR (%)	11.81	17.20	17.99	24.04	10.31
Fund Number	3	6	7.89	10	7.52
Year GP Established	1984	1996	1993.30	2005	14.70
Total Staff Count	31.50	92	1,566.44	300	7,607.70

Table 4: **Summary of fund performance and features.** The table shows the distribution of fund characteristics and performance in the sample.

	Final Return					
	Year 0-2	Year 3-4	Year 5+	Year 0-2	Year 3-4	Year 5+
	(1)	(2)	(3)	(4)	(5)	(6)
Sentiment	0.04***	0.02***	7.9×10^{-5}	0.03***	0.005	-0.003
	(0.006)	(0.006)	(0.006)	(0.01)	(0.009)	(0.009)
Interim Return	0.16***	0.62***	0.78***	0.16***	0.65***	0.72***
	(0.01)	(0.03)	(0.04)	(0.02)	(0.04)	(0.05)
Fixed-effects						
Fund				Yes	Yes	Yes
Fit statistics						
Observations	3,874	2,253	1,144	3,874	2,253	1,144
R^2	0.08	0.21	0.31	0.34	0.42	0.45
Within R^2				0.07	0.23	0.26
T-test: Sentiment Coef. Diff.		p = 0.00			p = 0.02	
T-test: Return Coef. Diff.		p = 0.00			p = 0.00	

Table 5: **Baseline analyses.** The regressions are run at the portfolio company - year (t) level. The sample is split by time since entry (Year 0-2, Year 3-4, Year 5+ since entry). Column (1) and Column (4) are run on the sample of observations that are in years 0-2 of an investment. Column (2) and Column (5) are run on the sample of observations that are in years 3-4 of an investment. Column (3) and Column (6) are run on the sample of observations that are in years 5 or later. Fund fixed effects are included in Columns (4), (5), and (6). Standard errors are clustered at the fund level in parentheses. There are no fixed effects for Columns (1), (2), and (3). Standard errors in parentheses are heteroskedasticity-robust. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. First t-test (under Column (2)) tests whether coefficients in Columns (1) and Column (3) are statistically different. Second t-test (under Column (5)) tests whether coefficients in Columns (4) and (6) are statistically different from each other. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return					
	Year 0-2		Year 3-4		Year 5+	
	(1)	(2)	(3)	(4)	(5)	(6)
Positive Sentiment	0.03***		0.01**		-0.005	
	(0.006)		(0.006)		(0.007)	
Interim Return	0.16***	0.17***	0.63***	0.62***	0.78***	0.77***
	(0.01)	(0.01)	(0.03)	(0.03)	(0.04)	(0.04)
Negative Sentiment		-0.02***		-0.01**		-0.006
		(0.007)		(0.007)		(0.005)
Fit statistics						
Observations	3,874	3,874	2,253	2,253	1,144	1,144
R^2	0.07	0.07	0.21	0.21	0.31	0.31
Adjusted R^2	0.07	0.07	0.21	0.21	0.31	0.31

Table 6: **Analyses splitting positive and negative sentiment.** The regressions are run at the portfolio company - year (t) level. The sample is split by time since entry (Year 0-2, Year 3-4, Year 5+ since entry). Column (1) and Column (2) are run on the sample of observations that are in years 0-2 of an investment. Column (3) and Column (4) are run on the sample of observations that are in years 3-4 of an investment. Column (5) and Column (6) are run on the sample of observations that are in years 5 or later. There are no fixed effects. Standard errors in parentheses are heteroskedasticity-robust. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Positive Sentiment is the number of positive words normalized by the overall number of words in the extracted text for a portfolio company in year t. Negative Sentiment is the number of negative words normalized by the overall number of words in the extracted text for a portfolio company in year t. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return					Last Interim Return	Final Return
	Year 0-2	Year 3-4	Year 5+	First Obs.	Last Obs.		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sentiment	0.04***	0.003	-0.008	0.02	0.006		
	(0.01)	(0.009)	(0.010)	(0.01)	(0.008)		
Interim Return	0.16***	0.64***	0.73***	0.14***	0.84***		
	(0.02)	(0.05)	(0.06)	(0.02)	(0.05)		
First Sentiment						0.04***	-0.01
						(0.008)	(0.009)
Last Sentiment							0.01
							(0.009)
Last Interim Return							0.84***
							(0.05)
Fixed-effects							
Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics							
Observations	1,427	1,586	877	1,102	1,102	1,102	1,102
R^2	0.36	0.44	0.44	0.31	0.51	0.30	0.51
Within R^2	0.09	0.23	0.22	0.07	0.34	0.02	0.34
T-test: Sentiment Coef. Diff.				p = 0.37			
T-test: Return Coef. Diff.				p = 0.00			

Table 7: **Analysis with repeat observations** The regressions are run at the portfolio company - year (t) level. All columns are run on the subsample where there are at least two observations that are a minimum of two years apart for a portfolio company. In Columns (1), (2), and (3) the sample is split by time since entry (Year 0-2, Year 3-4, Year 5+ since entry). Column (4) focuses on the first observation for each portfolio company in the subsample. Column (5) focuses on the last observation for each portfolio company in the subsample. We test whether the sentiment and return coefficients in Columns (4) and Column (5) are statistically different from each other. Columns (6) and (7) have one observation for each portfolio company in the sample. Fund fixed effects are used throughout. Standard errors are clustered at the fund level and are in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return		
	(1)	(2)	(3)
Sentiment	0.04***	0.03***	-0.003
	(0.006)	(0.009)	(0.003)
Interim Return	0.16***	0.16***	0.006
	(0.01)	(0.02)	(0.004)
Period = Year3-4	-0.07***	-0.07***	-0.002
	(0.01)	(0.01)	(0.002)
Period = Year5+	-0.08***	-0.06***	-0.002
	(0.01)	(0.01)	(0.003)
Sentiment × Period = Year3-4	-0.02***	-0.02**	0.003
	(0.009)	(0.01)	(0.002)
Sentiment × Period = Year5+	-0.04***	-0.04***	0.003
	(0.009)	(0.01)	(0.004)
Interim Return × Period = Year3-4	0.46***	0.43***	0.009
	(0.03)	(0.04)	(0.010)
Interim Return × Period = Year5+	0.62***	0.56***	-0.02
	(0.04)	(0.04)	(0.01)
Fixed-effects			
Fund		Yes	
Portfolio Company			Yes
Fit statistics			
Observations	7,271	7,271	7,271
R^2	0.14	0.34	0.99
Within R^2		0.13	0.004

Table 8: **Panel Specification** The regressions are run at the portfolio company - year (t) level using the complete sample. Time Since Entry is a categorical variable equal to either 0-2 years, 3-4 years, or 5+ years and groups observations by the year of an investment. The reference category is 0-2 years and is dropped due to collinearity. Column (1) has no fixed effects and standard errors in parentheses are heteroskedasticity-robust. Column (2) uses Fund fixed effects and standard errors are clustered at the fund level in parentheses. Column (3) uses Portfolio Company fixed effects and standard errors are clustered at the portfolio company level in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return					
	Year 0-2	Year 3-4	Year 5+	Year 0-2	Year 3-4	Year 5+
	(1)	(2)	(3)	(4)	(5)	(6)
Sentiment	0.04***	0.02***	0.003	0.03**	0.03	-0.02
	(0.007)	(0.006)	(0.008)	(0.02)	(0.02)	(0.02)
GP Size	0.03***	0.01*	0.01**			
	(0.005)	(0.006)	(0.006)			
Interim Return	0.16***	0.61***	0.76***	0.21***	0.57***	0.78***
	(0.01)	(0.03)	(0.05)	(0.03)	(0.07)	(0.10)
Sentiment × GP Size	0.008*	-8.6×10^{-5}	-0.002			
	(0.005)	(0.005)	(0.006)			
GP Size × Interim Return	-0.006	0.04*	-0.02			
	(0.01)	(0.02)	(0.02)			
GP Age				0.004***	0.002***	0.001*
				(0.0005)	(0.0006)	(0.0007)
Sentiment × GP Age				0.0004	-9.1×10^{-5}	0.001*
				(0.0005)	(0.0005)	(0.0006)
GP Age × Interim Return				-0.002*	0.0009	0.0002
				(0.0009)	(0.002)	(0.003)
Fit statistics						
Observations	3,436	1,945	940	3,466	1,966	954
R^2	0.09	0.20	0.31	0.09	0.21	0.31
Adjusted R^2	0.09	0.20	0.31	0.09	0.21	0.31

Table 9: **Heterogeneity across GPs** The regressions are run at the portfolio company - year (t) level. Columns (1), (2), and (3) regress final return on sentiment and interim return in year t, interacted with GP size. GP size is the log of the size of the fund normalized by prior five years' average global GP fundraising. Columns (4), (5), and (6) regress final return on sentiment and interim return in year t, interacted with the number of years a GP has existed as of 2020 (the end of our sample) (GP Age). The sample is split by time since entry (Year 0-2, Year 3-4, Year 5+ since entry). Column (1) and Column (4) are run on the sample of observations that are in years 0-2 of an investment. Column (2) and Column (5) are run on the sample of observations that are in years 3-4 of an investment. Column (3) and Column (6) are run on the sample of observations that are in years 5 or later. There are no GP fund fixed effects. Standard errors in parentheses are heteroskedasticity-robust. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return					
	Year 0-2	Year 3-4	Year 5+	Year 0-2	Year 3-4	Year 5+
	(1)	(2)	(3)	(4)	(5)	(6)
Post Round Ind.	0.005	-0.02	-0.02			
	(0.02)	(0.02)	(0.02)			
Sentiment	0.03***	0.02**	0.003	0.03***	0.008	-0.003
	(0.01)	(0.008)	(0.009)	(0.010)	(0.009)	(0.009)
Interim Return	0.17***	0.62***	0.73***	0.18***	0.64***	0.71***
	(0.02)	(0.05)	(0.08)	(0.02)	(0.04)	(0.05)
Post Round Ind. × Sentiment	-0.02	-0.04**	-0.01			
	(0.01)	(0.01)	(0.02)			
Post Round Ind. × Interim Return	-0.05	0.08	-0.006			
	(0.03)	(0.07)	(0.11)			
Round Year				-0.01	0.008	-0.10
				(0.02)	(0.06)	(0.14)
Round Year × Sentiment				-0.02	-0.07**	0.02
				(0.02)	(0.03)	(0.07)
Round Year × Interim Return				-0.09***	0.11	0.88
				(0.03)	(0.18)	(0.66)
Fixed-effects						
Fund	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	3,874	2,253	1,144	3,874	2,253	1,144
R^2	0.34	0.43	0.45	0.34	0.42	0.45
Within R^2	0.07	0.23	0.26	0.07	0.23	0.26

Table 10: **Heterogeneity by financing round.** The sample is split by time since entry (Year 0-2, Year 3-4, Year 5+ since entry). Columns (1) and (4) are run on the sample of observations that are in years 0-2 of an investment. Columns (2) and (5) are run on the sample of observations that are in years 3-4 of an investment. Columns (3) and (6) are run on the sample of observations that are in years 5 or later. Post Round Ind. is an indicator variable for whether the observation year is at any time after the portfolio company raised an additional round of funding. Round Year is an indicator variable for whether the observation occurs in the year of an additional round of financing. All columns use fund fixed effects and standard errors are clustered at the fund level in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company. There are 603 rounds in Year 0-2, 116 rounds in Year 3-4, and 39 rounds in Year 5+.

	Final Return		
	Year 0-2	Year 3-4	Year 5+
	(1)	(2)	(3)
Sentiment	0.04***	0.02***	0.003
	(0.007)	(0.006)	(0.006)
Interim Return	0.16***	0.61***	0.77***
	(0.01)	(0.03)	(0.04)
Operations Topic Ind.	-0.009	-0.02	-0.03
	(0.01)	(0.01)	(0.02)
Capital Structure Topic Ind.	-0.06***	0.006	0.02*
	(0.02)	(0.01)	(0.01)
Supply Chain and Customer Topic Ind.	-0.04***	-0.009	0.007
	(0.01)	(0.01)	(0.02)
Cash Flow Topic Ind.	0.05***	0.010	0.03*
	(0.01)	(0.01)	(0.02)
Macroeconomy Topic Ind.	-0.006	0.008	-0.02
	(0.02)	(0.01)	(0.02)
Leadership Topic Ind.	-0.004	-0.02	0.0003
	(0.01)	(0.02)	(0.02)
Business Model Topic Ind.	0.05***	0.03**	-0.02
	(0.01)	(0.01)	(0.02)
Fit statistics			
Observations	3,874	2,253	1,144
R^2	0.09	0.21	0.31
Adjusted R^2	0.09	0.21	0.31

Table 11: **Analyses with textual discussion topic controls.** The sample is split by time since entry (Year 0-2, Year 3-4, Year 5+ since entry). Column (1) is run on the sample of observations that are in years 0-2 of an investment. Column (2) is run on the sample of observations that are in years 3-4 of an investment. Column (3) is run on the sample of observations that are in years 5 or later. Operations Topic Ind., Capital Structure Topic Ind., Supply Chain and Customer Topic Ind., Cash Flow Topic Ind., Macroeconomy Topic Ind., Leadership Topic Ind., and Business Model Topic Ind. are all indicator variables equal to one if the corresponding topic is the most important (has the highest number of sentences) in the text of the portfolio company report for that year. There are no fixed effects. Standard errors in parentheses are heteroskedasticity-robust. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return						
	Operations	Capital Structure	Supply Chain and Customer	Cash Flow	Macroeconomy	Leadership	Business Model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sentiment	-0.004	0.01	-0.08*	0.12***	0.13***	0.07*	0.10***
	(0.04)	(0.04)	(0.05)	(0.02)	(0.05)	(0.04)	(0.04)
Interim Return	0.24***	0.14***	0.16***	0.16***	0.18***	0.11	0.13***
	(0.04)	(0.03)	(0.05)	(0.04)	(0.05)	(0.07)	(0.03)
Fixed-effects							
Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics							
Observations	661	779	603	640	480	590	834
R^2	0.57	0.54	0.58	0.50	0.49	0.53	0.46
Within R^2	0.14	0.03	0.09	0.11	0.11	0.05	0.06

Table 12: **Analyses of particular textual discussion topics (early observations only).** The regressions are run at the portfolio company - year (t) level. The sample consists only of observations that occur prior to year 3 of the investment. Each column is a subsample determined by which topic the portfolio company had as the main topic (most sentences) early in investment (years < 3). Column (1) is the subsample of portfolio companies that had Operations as the main topic (most sentences) early in the investment (years < 3). Column (2) is the subsample of portfolio companies that had Capital Structure as the main topic (most sentences) early in the investment (years < 3). Column (3) is the subsample of portfolio companies that had Supply Chain and Customer as the main topic (most sentences) early in the investment (years < 3). Column (4) is the subsample of portfolio companies that had Cash Flow as the main topic (most sentences) early in the investment (years < 3). Column (5) is the subsample of portfolio companies that had Macroeconomy as the main topic (most sentences) early in the investment (years < 3). Column (6) is the subsample of portfolio companies that had Leadership as the main topic (most sentences) early in the investment (years < 3). Column (7) is the subsample of portfolio companies that had Business Model as the main topic (most sentences) early in the investment (years < 3). Fund fixed effects are used throughout. Standard errors are clustered at the fund level and are in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words using only sentences about that column's topic from the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return						
	Operations	Capital Structure	Supply Chain and Customer	Cash Flow	Macroeconomy	Leadership	Business Model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sentiment	-0.002	0.01	-0.02	0.01	0.01	-0.009	0.02
	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
Interim Return	0.62***	0.41***	0.36***	0.59***	0.58***	0.46***	0.57***
	(0.09)	(0.09)	(0.09)	(0.09)	(0.10)	(0.08)	(0.07)
Early Observation	0.01	0.03	0.005	0.07***	0.04	0.04**	0.06***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Sentiment × Early Observation	0.05**	0.005	0.02	0.03**	0.05***	0.02	0.03*
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
Interim Return × Early Observation	-0.42***	-0.28***	-0.21**	-0.42***	-0.45***	-0.34***	-0.43***
	(0.09)	(0.08)	(0.09)	(0.08)	(0.09)	(0.09)	(0.06)
Fixed-effects							
Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics							
Observations	1,379	1,548	1,185	1,424	991	1,208	1,666
R^2	0.57	0.53	0.56	0.51	0.46	0.52	0.45
Within R^2	0.16	0.05	0.08	0.15	0.12	0.07	0.11

Table 13: **Analyses of particular textual discussion topics (all observations).** The regressions are run at the portfolio company - year (t) level. Early Observation is an indicator variable for whether the observation occurs prior to year 3 of the investment. Each column is a subsample determined by which topic the portfolio company had as the main topic (most sentences) early in investment (years < 3). We use the complete set of observations for each portfolio company. Column (1) is the subsample of portfolio companies that had Operations as the main topic (most sentences) early in the investment (years < 3). Column (2) is the subsample of portfolio companies that had Capital Structure as the main topic (most sentences) early in the investment (years < 3). Column (3) is the subsample of portfolio companies that had Supply Chain and Customer as the main topic (most sentences) early in the investment (years < 3). Column (4) is the subsample of portfolio companies that had Cash Flow as the main topic (most sentences) early in the investment (years < 3). Column (5) is the subsample of portfolio companies that had Macroeconomy as the main topic (most sentences) early in the investment (years < 3). Column (6) is the subsample of portfolio companies that had Leadership as the main topic (most sentences) early in the investment (years < 3). Column (7) is the subsample of portfolio companies that had Business Model as the main topic (most sentences) early in the investment (years < 3). Fund fixed effects are used throughout. Standard errors are clustered at the fund level and are in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

Appendix

	Final Return		
	Year 0-2	Year 3-4	Year 5+
	(1)	(2)	(3)
Sentiment	0.04***	0.007	-0.005
	(0.010)	(0.009)	(0.010)
Interim Return	0.16***	0.69***	0.71***
	(0.02)	(0.04)	(0.06)
Sentiment × VC Fund Ind.	-0.05	-0.02	0.02
	(0.04)	(0.03)	(0.03)
VC Fund Ind. × Interim Return	-0.05	-0.27**	0.05
	(0.04)	(0.12)	(0.07)
Fixed-effects			
Fund	Yes	Yes	Yes
Fit statistics			
Observations	3,874	2,253	1,144
R^2	0.34	0.43	0.45
Within R^2	0.07	0.23	0.26

Table A1: **Analysis with interaction for VC investors.** The regressions are run at the portfolio company - year (t) level. VC Fund Ind. is an indicator variable equal to 1 if the GP is a venture capital investor. The sample is split by time since entry (Year 0-2, Year 3-4, Year 5+ since entry). Column (1) is run on the sample of observations that are in years 0-2 of an investment. Column (2) is run on the sample of observations that are in years 3-4 of an investment. Column (3) is run on the sample of observations that are in years 5 or later. Fund fixed effects are used throughout. Standard errors are clustered at the fund level in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return					
	Year 0-2	Year 3-4	Year 5+	Year 0-2	Year 3-4	Year 5+
	(1)	(2)	(3)	(4)	(5)	(6)
Sentiment	0.04***	0.02***	-0.003	0.04***	0.007	-0.005
	(0.006)	(0.006)	(0.006)	(0.010)	(0.009)	(0.010)
Interim Return	0.15***	0.66***	0.77***	0.16***	0.69***	0.71***
	(0.01)	(0.03)	(0.05)	(0.02)	(0.04)	(0.06)
Fixed-effects						
Fund				Yes	Yes	Yes
Fit statistics						
Observations	3,309	1,939	991	3,309	1,939	991
R^2	0.09	0.28	0.30	0.31	0.46	0.47
Within R^2				0.08	0.28	0.25

Table A2: **Subsample dropping all VC observations: Sample split by time since entry.** The regressions are run at the portfolio company - year (t) level. The sample is split by time since entry (Year 0-2, Year 3-4, Year 5+ since entry). Column (1) and Column (4) are run on the sample of observations that are in years 0-2 of an investment. Column (2) and Column (5) are run on the sample of observations that are in years 3-4 of an investment. Column (3) and Column (6) are run on the sample of observations that are in years 5 or later. Fund fixed effects are included in Columns (4), (5), and (6). Standard errors are clustered at the fund level in parentheses. There are no fixed effects for Columns (1), (2), and (3). Standard errors in parentheses are heteroskedasticity-robust. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return		
	Year 0-2	Year 3-4	Year 5+
	(1)	(2)	(3)
Sentiment	0.04***	0.02***	-0.003
	(0.007)	(0.006)	(0.006)
Interim Return	0.15***	0.66***	0.76***
	(0.01)	(0.03)	(0.05)
Operations Topic Ind.	-0.02	-0.03*	-0.03
	(0.01)	(0.01)	(0.02)
Capital Structure Topic Ind.	0.04**	0.03**	0.01
	(0.01)	(0.01)	(0.01)
Supply Chain and Customer Topic Ind.	-0.02	0.003	0.003
	(0.01)	(0.01)	(0.02)
Cash Flow Topic Ind.	0.03**	0.01	0.04**
	(0.01)	(0.01)	(0.02)
Macroeconomy Topic Ind.	-0.03*	0.009	-0.04**
	(0.02)	(0.01)	(0.02)
Leadership Topic Ind.	-0.004	-0.02	0.001
	(0.01)	(0.02)	(0.02)
Business Model Topic Ind.	0.04***	0.01	-0.002
	(0.01)	(0.01)	(0.02)
Fit statistics			
Observations	3,309	1,939	991
R^2	0.09	0.28	0.31
Adjusted R^2	0.09	0.28	0.30

Table A3: **Subsample dropping all VC observations: Topic Analysis.** The sample is split by time since entry (Year 0-2, Year 3-4, Year 5+ since entry). Column (1) is run on the sample of observations that are in years 0-2 of an investment. Column (2) is run on the sample of observations that are in years 3-4 of an investment. Column (3) is run on the sample of observations that are in years 5 or later. Operations Topic Ind., Capital Structure Topic Ind., Supply Chain and Customer Topic Ind., Cash Flow Topic Ind., Macroeconomy Topic Ind., Leadership Topic Ind., and Business Model Topic Ind. are all indicator variables equal to one if the corresponding topic is the most important (has the highest number of sentences) in the text of the portfolio company report for that year. There are no fixed effects. Standard errors in parentheses are heteroskedasticity-robust. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return						
	Operations	Capital Structure	Supply Chain and Customer	Cash Flow	Macroeconomy	Leadership	Business Model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sentiment	-0.002	0.01	-0.01	0.01	0.01	-0.007	0.03**
	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
Interim Return	0.63***	0.49***	0.35***	0.61***	0.57***	0.47***	0.58***
	(0.10)	(0.06)	(0.10)	(0.09)	(0.10)	(0.09)	(0.07)
Early Observation	0.02	0.05***	0.01	0.07***	0.04	0.04*	0.07***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Sentiment × Early Observation	0.05**	0.02	0.03	0.04**	0.05***	0.02	0.03*
	(0.03)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Interim Return × Early Observation	-0.42***	-0.36***	-0.19**	-0.44***	-0.45***	-0.34***	-0.45***
	(0.10)	(0.06)	(0.09)	(0.08)	(0.10)	(0.09)	(0.06)
Fixed-effects							
Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics							
Observations	1,268	1,246	1,079	1,341	918	1,059	1,565
R^2	0.56	0.52	0.57	0.50	0.44	0.51	0.44
Within R^2	0.16	0.10	0.09	0.16	0.11	0.09	0.12

Table A4: **Subsample dropping all VC observations: Topic by early observation analysis.** The regressions are run at the portfolio company - year (t) level. Early Observation is an indicator variable for whether the observation occurs prior to year 3 of the investment. Each column is a subsample determined by which topic the portfolio company had as the main topic (most sentences) early in investment (years < 3). We use the complete set of observations for each portfolio company. Fund fixed effects are used throughout. Standard errors are clustered at the fund level and are in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return					
	Year 0-2	Year 3-4	Year 5+	Year 0-2	Year 3-4	Year 5+
	(1)	(2)	(3)	(4)	(5)	(6)
Missing	-0.02	-0.008	-0.03	-0.03	-0.03	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Fixed-effects						
Fund				Yes	Yes	Yes
Fit statistics						
Observations	2,932	1,956	892	2,932	1,956	892
R^2	0.0002	0.0001	0.002	0.27	0.26	0.28
Within R^2				0.0009	0.001	0.003

Table A5: **Analysis of missing observations on full sample: Sample split by years since investment.** The regressions are run at the portfolio company level (unlike Table 5, which is run at portfolio company - year). Missing is an indicator for whether we have a report from a GP within the investment window for the portfolio company, but do not have an observation for the portfolio company. In Columns (1) and (4), missing is equal to 1 for a portfolio company if any of the observations are missing in years 0-2 of investment. In Columns (2) and (5), missing is equal to 1 for a portfolio company if any of the observations are missing in years 3-4 of investment. In Columns (3) and (6), missing is equal to 1 for a portfolio company if any of the observations are missing in year 5 or later after the investment. Standard errors are clustered at the fund level and are in parentheses for Columns (4), (5), and (6). Heteroskedasticity-robust standard errors are used in Columns (1), (2), and (3). *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return					
	Year 0-2	Year 3-4	Year 5+	Year 0-2	Year 3-4	Year 5+
	(1)	(2)	(3)	(4)	(5)	(6)
Missing	-0.007	-0.009	-0.02	-0.03	-0.03	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Fixed-effects						
Fund				Yes	Yes	Yes
Fit statistics						
Observations	2,617	1,892	846	2,617	1,892	846
R^2	5.9×10^{-5}	0.0001	0.001	0.26	0.26	0.27
Within R^2				0.0009	0.001	0.003

Table A6: **Analysis of missing observations on subsample of funds with at least two reports.** This table repeats the previous table, now restricting the sample to funds with at least two reports. The sample is split by years since investment. The regressions are run at the portfolio company level (unlike Table 5, which is run at portfolio company - year). Missing is an indicator for whether we have a report from a GP within the investment window for the portfolio company, but do not have an observation for the portfolio company. In Columns (1) and (4), missing is equal to 1 for a portfolio company if any of the observations are missing in years 0-2 of investment. In Columns (2) and (5), missing is equal to 1 for a portfolio company if any of the observations are missing in years 3-4 of investment. In Columns (3) and (6), missing is equal to 1 for a portfolio company if any of the observations are missing in year 5 or later after the investment. Standard errors are clustered at the fund level and are in parentheses for Columns (4), (5), and (6). Heteroskedasticity-robust standard errors are used in Columns (1), (2), and (3). *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Next Year Missing					
	Year 0-2	Year 3-4	Year 5+	Year 0-2	Year 3-4	Year 5+
	(1)	(2)	(3)	(4)	(5)	(6)
Interim Return	-0.01		-0.007	-0.02		-0.02
	(0.01)		(0.01)	(0.01)		(0.01)
Sentiment		-0.01**	-0.01*		-0.002	-0.001
		(0.006)	(0.007)		(0.007)	(0.007)
Fixed-effects						
Fund				Yes	Yes	Yes
Fit statistics						
Observations	4,316	4,316	4,316	4,316	4,316	4,316
R^2	0.0002	0.001	0.001	0.26	0.26	0.26
Within R^2				0.0005	1.6×10^{-5}	0.0005

Table A7: **Predictors of a portfolio company missing a report next year.** The regressions are run at the portfolio company - year (t) level. The sample is restricted to funds with at least two reports and to all observations where we have Interim Return. Next Year Missing is an indicator for whether in the following year we have a report from a GP within the investment window for the portfolio company, but do not have an observation for the portfolio company. Standard errors are clustered at the fund level and are in parentheses for Columns (4), (5), and (6). Heteroskedasticity-robust standard errors are used in Columns (1), (2), and (3). *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit.

	Next Year Missing (1)	Any Year Missing (2)	Two Years Later Missing (3)	Next Year Missing (4)	Any Year Missing (5)	Two Years Later Missing (6)
Interim Return	-0.007	-0.01	0.008	-0.004	0.0003	0.01
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.03)
Sentiment	-0.02**	-0.02**	-0.02*	-0.004	0.0008	-0.02
	(0.009)	(0.009)	(0.01)	(0.01)	(0.01)	(0.01)
Fixed-effects						
Fund				Yes	Yes	Yes
Fit statistics						
Observations	2,115	2,752	1,232	2,115	2,752	1,232
R^2	0.002	0.002	0.003	0.41	0.34	0.43
Within R^2				0.0001	3.9×10^{-6}	0.003

Table A8: **Predicting missing observations from portfolio companies' first observation.** The regressions are run at the portfolio company level (collapsing the sample to a single observation per portfolio company). The sample is restricted to funds with at least two reports. Interim Return is the first Interim Return for the portfolio company in our sample. Sentiment is the sentiment in the same year. Next Year Missing is an indicator for whether in the following year we have a report from a GP within the investment window for the portfolio company, but do not have an observation for the portfolio company. Any Year Missing is an indicator for whether in any following year we have a report from a GP within the investment window for the portfolio company, but do not have an observation for the portfolio company. Two Years Later Missing is an indicator for whether two years later we have a report from a GP within the investment window for the portfolio company, but do not have an observation for the portfolio company. Standard errors are clustered at the fund level and are in parentheses for Columns (4), (5), and (6). Heteroskedasticity-robust standard errors are used in Columns (1), (2), and (3). *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit.

	Next Year Missing (1)	Any Year Missing (2)	Two Years Later Missing (3)	Next Year Missing (4)	Any Year Missing (5)	Two Years Later Missing (6)
Interim Return	-0.003	-0.008	0.01	-0.002	0.0007	0.02
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)
Sentiment	-0.02**	-0.02**	-0.03**	-0.01	-0.008	-0.04**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Fixed-effects						
Fund				Yes	Yes	Yes
Fit statistics						
Observations	1,723	2,268	981	1,723	2,268	981
R^2	0.003	0.003	0.005	0.41	0.35	0.47
Within R^2				0.0007	0.0003	0.008

Table A9: **Predicting missing observations from portfolio companies' first early observation.** This table repeats the previous table, but retains only portfolio companies where the first observation is early (years 0-2) after the investment. The regressions are run at the portfolio company level (collapsing the sample down to a single observation per portfolio company). The sample is restricted to funds with at least two reports. Interim Return is the first Interim Return for the portfolio company in our sample. Sentiment is the sentiment in the same year. Next Year Missing is an indicator for whether in the following year we have a report from a GP within the investment window for the portfolio company, but do not have an observation for the portfolio company. Any Year Missing is an indicator for whether in any following year we have a report from a GP within the investment window for the portfolio company, but do not have an observation for the portfolio company. Two Years Later Missing is an indicator for whether two years later we have a report from a GP within the investment window for the portfolio company, but do not have an observation for the portfolio company. Standard errors are clustered at the fund level and are in parentheses for Columns (4), (5), and (6). Heteroskedasticity-robust standard errors are used in Columns (1), (2), and (3). *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit.

	Final Return					
	(1)	(2)	(3)	(4)	(5)	(6)
Sentiment	High	Low	High	Low	High	Low
Time Since Entry	Year 0-2		Year 3-4		Year 5+	
Sentiment	0.007	0.03	0.004	0.007	-0.009	-0.01
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Interim Return	0.16***	0.16***	0.59***	0.63***	0.79***	0.76***
	(0.02)	(0.02)	(0.04)	(0.05)	(0.06)	(0.06)
Fit statistics						
Observations	1,937	1,937	1,126	1,127	572	572
R^2	0.07	0.05	0.19	0.21	0.30	0.30
Adjusted R^2	0.07	0.05	0.18	0.21	0.30	0.30

Table A10: **Analyses split by above and below median sentiment.** The regressions are run at the portfolio company - year (t) level. The sample is split by time since entry (Year 0-2, Year 3-4, Year 5+ since entry), and by whether sentiment is above or below median within the group of years of the investment. Column (1) and Column (2) are run on the sample of observations that are in years 0-2 of an investment. Column (3) and Column (4) are run on the sample of observations that are in years 3-4 of an investment. Column (5) and Column (6) are run on the sample of observations that are in years 5 or later. For the second sample split, Columns (1), (3), and (5) are run on the subsample with above median (high) sentiment for that investment year group and Column (2), (4), and (6) are run on the subsample with below median (low) sentiment for each investment year group. There are no fixed effects. Standard errors in parentheses are heteroskedasticity-robust. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return					
	Year 0-2	Year 3-4	Year 5+	Year 0-2	Year 3-4	Year 5+
	(1)	(2)	(3)	(4)	(5)	(6)
Sentiment	0.04***	-0.0006	-0.005	0.04***	-0.0003	-0.005
	(0.01)	(0.01)	(0.010)	(0.01)	(0.01)	(0.010)
Interim Return	0.14***	0.69***	0.72***	0.27***	0.58***	0.36
	(0.03)	(0.05)	(0.06)	(0.10)	(0.16)	(0.22)
Update	0.08***	0.05	-0.003	0.09***	0.03	-0.02
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Interim Return × Update				-0.14	0.12	0.39
				(0.10)	(0.15)	(0.24)
Fixed-effects						
Fund	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	2,536	1,654	937	2,536	1,654	937
R^2	0.35	0.47	0.42	0.35	0.47	0.42
Within R^2	0.06	0.25	0.22	0.06	0.25	0.23

Table A11: **Analysis with portfolio company valuation update.** The regressions are run at the portfolio company - year (t) level. Update is an indicator variable equal to one if the valuation in year t changed by more than one percent from the prior year. Sample sizes are smaller because observations with no data in the preceding year are dropped. The sample is split by time since entry (Year 0-2, Year 3-4, Year 5+ since entry). Column (1) and Column (4) are run on the sample of observations that are in years 0-2 of an investment. Column (2) and Column (5) are run on the sample of observations that are in years 3-4 of an investment. Column (3) and Column (6) are run on the sample of observations that are in years 5 or later. Fund fixed effects are used throughout. Standard errors are clustered at the fund level and are in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return			
	(1)	(2)	(3)	(4)
Sentiment	0.06***	0.05***	0.06***	0.05***
	(0.010)	(0.01)	(0.008)	(0.01)
Time Since Entry Group	-0.05***	-0.04***	-0.05***	-0.06***
	(0.005)	(0.008)	(0.010)	(0.01)
Interim Return	-0.19***	-0.16***	-0.19***	-0.16***
	(0.03)	(0.03)	(0.04)	(0.04)
Sentiment × Time Since Entry Group	-0.02***	-0.02***	-0.02***	-0.02***
	(0.005)	(0.006)	(0.005)	(0.004)
Time Since Entry Group × Interim Return	0.36***	0.33***	0.36***	0.33***
	(0.02)	(0.02)	(0.02)	(0.02)
Fixed-effects				
Fund		Yes		Yes
Year			Yes	Yes
Fit statistics				
Observations	7,271	7,271	7,271	7,271
R^2	0.14	0.34	0.15	0.35
Within R^2		0.13	0.14	0.13

Table A12: **Analysis with time since entry interaction.** The regressions are run at the portfolio company - year (t) level. Time Since Entry Group is equal to 1 if the observation occurs 0-2 years after the investment, equal to 2 if the observation is 3-4 years after the investment, or equal to 3 if the observation is 5 or more years after the investment. Column (1) has no fixed effects. Standard errors in parentheses are heteroskedasticity-robust. Column (2) uses fund fixed effects and standard errors are clustered at the fund level in parentheses. Column (3) uses year fixed effects and standard errors are clustered at the year level in parentheses. Column (4) uses year and fund fixed effects and standard errors are clustered at the year and fund level in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return		
	Winsorize top 5%	Winsorize if <0% Return	Winsorize if (1) or (2)
	(1)	(2)	(3)
Sentiment	0.03***	0.02***	0.02***
	(0.01)	(0.004)	(0.004)
Interim Return	0.14***	0.11***	0.09***
	(0.02)	(0.01)	(0.009)
Fixed-effects			
Fund	Yes	Yes	Yes
Fit statistics			
Observations	3,874	3,874	3,874
R^2	0.34	0.33	0.34
Within R^2	0.06	0.09	0.09

Table A13: **Analyses with winsorization.** The regressions are run at the portfolio company - year (t) level. Columns (1), (2), and (3) are run on the subsample of observations that are in years 0-2 of an investment. Column (1) uses data where the top 5% of final return has been winsorized. Column (2) uses data where returns below 0% have been winsorized. Column (3) uses data where both the top 5% of final returns and final returns below 0% have been winsorized. Fund fixed effects are included in Columns (1), (2), and (3). Standard errors are clustered at the fund level in parentheses. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return		
	Year 0-2	Year 3-4	Year 5+
	(1)	(2)	(3)
Sentiment	0.04**	0.03*	0.005
	(0.01)	(0.01)	(0.01)
Sector R&D Intensity	0.94***	0.20	-0.17
	(0.30)	(0.40)	(0.62)
Interim Return	0.14***	0.53***	0.78***
	(0.03)	(0.07)	(0.11)
Sentiment × Sector R&D Intensity	-0.11	-0.42	-0.18
	(0.23)	(0.27)	(0.32)
Sector R&D Intensity × Interim Return	0.22	2.1*	-1.1
	(0.56)	(1.1)	(1.8)
Fixed-effects			
Fund	Yes	Yes	Yes
Fit statistics			
Observations	3,873	2,253	1,144
R^2	0.34	0.43	0.45
Within R^2	0.07	0.23	0.26

Table A14: **Analysis with interaction for sector R&D intensity.** The regressions are run at the portfolio company - year (t) level. Sector R&D Intensity is the R&D intensity of the portfolio company's two-digit GICS sector, calculated using the average ratio of R&D expense to revenue of public companies in Compustat in that sector between 2005 and 2015. The sample is split by time since entry (Year 0-2, Year 3-4, Year 5+ since entry). Column (1) is run on the sample of observations that are in years 0-2 of an investment. Column (2) is run on the sample of observations that are in years 3-4 of an investment. Column (3) is run on the sample of observations that are in years 5 or later. Fund fixed effects are included in Columns (1), (2), and (3). Standard errors are clustered at the fund level in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return				
	Year 1	Year 2	Year 3	Year 4	Year 5+
	(1)	(2)	(3)	(4)	(5)
Sentiment	0.04***	0.04***	0.02***	0.010	7.9×10^{-5}
	(0.009)	(0.009)	(0.008)	(0.009)	(0.006)
Interim Return	0.12***	0.39***	0.57***	0.71***	0.78***
	(0.01)	(0.03)	(0.04)	(0.05)	(0.04)
Fit statistics					
Observations	2,082	1,792	1,328	925	1,144
R^2	0.06	0.14	0.19	0.26	0.31
Adjusted R^2	0.06	0.14	0.19	0.25	0.31

Table A15: **Analysis with sample split by year.** The regressions are run at the portfolio company - year (t) level. The sample is split by years since entry (instead of grouping years 1 through 5). There are no fixed effects. Standard errors in parentheses are heteroskedasticity-robust. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1. Interim Return is the annualized return for an investor from the initial entry value for a portfolio company to the portfolio company value at a given observation prior to exit. Final Return is the annualized return for the investor from the initial entry value for a portfolio company to the final exit value for a portfolio company.

	Final Return					
	Year 0-2	Year 3-4	Year 5+	Year 0-2	Year 3-4	Year 5+
	(1)	(2)	(3)	(4)	(5)	(6)
Sentiment	0.15***	0.10**	0.04	0.15***	0.06	0.06
	(0.03)	(0.04)	(0.04)	(0.04)	(0.07)	(0.07)
Interim Return	0.79***	1.1***	1.3***	0.83***	1.2***	1.2***
	(0.05)	(0.04)	(0.05)	(0.07)	(0.06)	(0.07)
Fixed-effects						
Fund				Yes	Yes	Yes
Fit statistics						
Observations	3,874	2,253	1,144	3,874	2,253	1,144
R^2	0.12	0.30	0.45	0.34	0.48	0.63
Within R^2				0.12	0.34	0.43

Table A16: **Analysis with returns that are not annualized.** The regressions are run at the portfolio company - year (t) level. The sample is split by time since entry (Year 0-2, Year 3-4, Year 5+ since entry). Column (1) and Column (4) are run on the sample of observations that are in years 0-2 of an investment. Column (2) and Column (5) are run on the sample of observations that are in years 3-4 of an investment. Column (3) and Column (6) are run on the sample of observations that are in years 5 or later. Fund fixed effects are included in Columns (4), (5), and (6). Standard errors are clustered at the fund level in parentheses. There are no fixed effects for Columns (1), (2), and (3). Standard errors in parentheses are heteroskedasticity-robust. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1.

	Sentiment						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Operations Topic Change	0.48***						
	(0.09)						
Capital Structure Topic Change		-0.37***					
		(0.09)					
Supply Chain and Customer Topic Change			0.05				
			(0.09)				
Cash Flow Topic Change				0.02			
				(0.06)			
Macroeconomy Topic Change					-0.37***		
					(0.09)		
Leadership Topic Change						-0.11	
						(0.08)	
Business Model Topic Change							0.30***
							(0.09)
Fixed-effects							
Fund	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics							
Observations	3,301	3,301	3,301	3,301	3,301	3,301	3,301
R^2	0.35	0.35	0.35	0.35	0.35	0.35	0.35
Within R^2	0.01	0.007	7.3×10^{-5}	2.3×10^{-5}	0.006	0.0005	0.004

Table A17: **Analysis of how changes in the intensity of a topic's discussion correlate with sentiment.** The regressions are run at the portfolio company - year (t) level. Each Topic Change measure is calculated as (Ratio of sentences about topic at t) - (Ratio of sentences about topic at $t-1$). Standard errors are clustered at the fund level in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively. Sentiment is the number of positive words minus the number of negative words and then normalized by the overall number of words in the extracted text for a portfolio company in year t. We z-score normalize across all portfolio company years so the distribution of sentiment scores has mean zero and a standard deviation of 1.