Using Machine Learning to Estimate the Effect of General Partners on Venture Capital Performance

Abstract

Despite extensive research on performance persistence in the venture capital (VC) industry, evidence supporting the common assumption that general partners drive such persistence is limited. In this paper, we employ machine learning methods to quantify the effect (if any) of general partners (GPs) on performance persistence across multiple VC funds. Analyzing 29,021 quarterly observations from 722 funds managed by 811 general partners between 1997 and 2022, we document statistically significant, albeit modest, effects of GPs on performance persistence. These magnitudes are substantially smaller than those found in the literature, highlighting the limited external validity of traditional methods. We also find that GP effects consistently exceed VC firm effects, suggesting that individual-level analyses provide greater insights than firm-level ones. Broadly, our results suggest that most of the variation in VC performance is not attributable to the organizational characteristics of VC firms and can be only partially explained by the individuals managing them.

Keywords: *Venture capital; General partners; Machine learning; Performance persistence*

1. Introduction

Venture capital (VC) firms are a critical driver of innovation and economic growth (Gompers and Lerner 2001, Lerner and Nanda 2020) and, over the past years, have committed an impressive amount of resources to new ventures. In the US alone, VC firms closed 14,320 deals totaling $215.4 billion in 2024.[1] Organizationally, these investment activities are conducted by means of specific funds led by general partners (GPs). GPs are responsible for managing the capital committed by limited partners (LPs), namely institutional investors such as pension funds, endowments and family offices, and for making investment decisions in startups with high growth potential. GPs are tasked with sourcing, screening and selecting companies and, upon investment, monitoring them and orchestrating exit strategies to maximize financial returns (Gorman and Sahlman 1989, Tyebjee and Bruno 1984). Notoriously, these activities involve substantial judgment under conditions of high uncertainty (Ewens et al. 2018), information asymmetry (Amit et al. 1998, Cumming and Johan 2008) and right-skewed outcomes (Korteweg and Nagel 2016) given that a significant number of startups fail and generous VC fund returns are usually generated by a few successful investments. Moreover, investments are typically highly illiquid (Lerner and Schoar 2004) and outcomes unfold over long time horizons – typically spanning 7-10 years – complicating performance assessment and the detection of successful investment decisions.

These features raise the important questions of how well VC funds perform and whether their ability to generate financial returns can be sustained over time. These questions hold relevance in the strategy literature, which has long been concerned with the persistence of firms' competitive advantage (e.g., Black and Boal 1994). Evidence indicates that VC firms differ from other forms of actively managed funds by exhibiting *persistence* in financial performance over time (Kaplan and Schoar 2005, Cong and Xiao 2022, Nanda et al. 2020, Ewens and Rhodes-Kropf 2015, Hochberg et al. 2013). Such persistence has proven to be remarkably consistent over the last three decades, whereas other private funds, such as buyout funds, have experienced a decline in persistence (Harris et al. 2023).

[1] Source: https://nvca.org/press_releases/nvca-releases-2025-yearbook-showcasing-2024-vc-trends (consulted online on November 29, 2025).

Scholars in strategy and finance have offered various explanations for the origins of performance persistence (Fitza et al. 2009, Korteweg and Sorensen 2017). Some argue that GPs possess heterogeneous levels of skills and human capital, with the more adept consistently outperforming their peers (e.g., Ewens and Rhodes-Kropf 2015, Hochberg et al. 2013). Others vouch for the role of luck (e.g., Cong and Xiao 2022), i.e., the positive occurrence of investing in the right place at the right time, which endows GPs with track record, reputation, network and, consequently, a higher-quality deal flow and better investment opportunities in subsequent funds. All these arguments, even those emphasizing the role of luck, ultimately rest on the premise that GPs are instrumental in driving fund outcomes. According to Nanda et al. (2020), persistence can emerge not because some GPs are inherently more skilled but because initial luck sets in motion a path-dependent process whereby GPs, once successful, gain structural advantages that translate into sustained performance. The view that individual GPs may drive performance persistence (i.e., a time-invariant performance premium) rests on the implicit assumption that GPs can affect fund performance at all, i.e., that *managers matter*, echoing parallel research on the CEOs of business firms (Bertrand and Schoar 2003, Bennedsen et al. 2020). While this assumption may seem trivial, other streams of research show small or insignificant CEO effects (Shimao et al. 2025, Fee et al. 2013, Jarosiewicz and Ross 2023), casting doubt on whether top managers have idiosyncratic styles and whether they really influence firm performance.

In this paper, we empirically test the assumption that GPs matter. Uncovering a persistent effect of GPs on performance would support the view that GPs, whether skilled or lucky, play a central role in driving fund outcomes, whereas the absence of such an effect would raise further questions about the sources of persistence in VC. Given that performance persistence may also be driven by firm-level characteristics (e.g., Fitza et al. 2009), we empirically disentangle the effect of VC firms from that of individual GPs.

Prior studies in this research domain (e.g., Fitza et al. 2009, Ewens and Rhodes-Kropf 2015) have provided valuable evidence using in-sample variance decomposition or fixed effects (FE) methods. Yet, recent evidence (Shimao et al. 2025) shows that these methods are prone to overfitting when applied to high-dimensional sparse data, and the resulting evidence has been challenged

(Jarosiewicz and Ross 2023).[2] Similar to Shimao et al. (2025), we employ fourteen distinct machine learning (ML) algorithms, from regularized linear models to ensemble methods and kernel machines. These algorithms are specifically designed to handle overfitting, contrary to linear models used in traditional variance decomposition analyses. We then apply time-series cross-validation with expanding windows to avoid look-ahead bias (Bergmeir and Benítez 2012) to the time series of fund-level internal rate of return (IRR) for model selection. Using ML techniques to discern the sources of performance persistence is a novel and relevant endeavor that, as we will demonstrate, yields new results for the VC literature.

Our sample is a large-scale panel dataset from PitchBook's universe of VC funds, initially tracking 13,091 GPs across 14,116 funds with detailed fund-GP affiliation matches and quarterly fund-level IRR from 1997 to 2022. We apply a number of criteria to this broad sample to detect a potential GP-level effect on performance persistence: a minimum of two funds managed and at least 40 quarters observed (i.e., the typical lifespan of a VC fund) for the same individual. In doing so, we isolate a high-quality sub-sample of 811 GPs who managed 722 funds in 217 VC firms, comprising 29,021 VC fund-quarter observations.

For the analysis, we compare three model specifications across all ML algorithms: (1) a baseline model that includes fund-specific controls (e.g., vintage year, LPs' capital commitment) and equity market returns; (2) a GP-enhanced model with GP-level fixed effects; and (3) a VC-enhanced model with VC firm-level fixed effects. As in Shimao et al. (2025), we gauge the overall performance of the different model specifications by using the percentage improvement in out-of-sample R^2. We assess the statistical significance of these effects using paired t-tests on the R^2 series from the cross-validation procedure, evaluating the magnitude of improvement over the baseline model. To account for temporal dependence and to avoid normality assumptions, we also implement a wild bootstrap version of paired

[2] Overfitting occurs when statistical models capture noise or random variations specific to the estimation sample rather than capturing the true underlying relationships. This leads to inflated in-sample R^2 that fails to generalize to new data. Overfitting is particularly concerning when establishing GP or, more generally, managerial effects because in-sample differences between GPs may reflect sample-specific noise rather than a managerial effect that would persist in future funds.

t-tests (Davidson and Flachaire 2008).[3] Finally, we apply Benjamini and Hochberg's (1995) correction for multiple testing across different ML algorithms.

We obtain a number of novel results. First, we demonstrate that linear models severely overfit when used to analyze the persistent role of GPs. Applying ridge with a tiny penalization[4] to our data produces an in-sample R^2 of 74.1%, which is comparable to estimates in prior studies (see Fitza et al. 2009), but an out-of-sample R^2 of 43.5%, revealing a 30.6% overfitting (see Table 1). Second, we document a significant improvement in out-of-sample R^2 across all ML algorithms when including GP fixed effects. This result supports the assumption of a persistent GP effect. Third, we focus on the best-performing ML algorithm (i.e., the random forest), which explains 74.2% of the variance in fund performance in the baseline model. The inclusion of GP fixed effects in this model raises the explained variance to 77.8% – an improvement of 3.6 percentage points – suggesting a significant effect of GPs on performance. Yet, compared to prior estimates (e.g., Ewens and Rhodes-Kropf 2015, which documented a ~47% GP effect), our estimate is much smaller in magnitude.[5] We argue that this difference reflects overfitting inherent in in-sample fixed effect approaches and that the 3.6% effect represents a more accurate estimate of the persistent GP effect. In the literature on CEOs, magnitudes vary between around 10% and 30% (e.g., Crossland and Hambrick 2007, Lieberson and O'Connor 1972, Mackey 2008, Wasserman et al. 2001), while, more recently, Shimao et al. (2025) found an insignificant effect.

Insert Table 1 about here

----

[3] The results of our wild bootstrap t-tests are consistent with traditional paired t-tests. Both methods yield p-values well below 0.001 for the GP effects, suggesting that our findings are not artifacts of any particular statistical inference method.

[4] In our case, the standard OLS matrix inversion fails completely and a ridge regression with minimal penalization approximates OLS behavior while handling the computational singularity.

[5] A direct comparison of in-sample and out-of-sample estimates is not straightforward. While in-sample estimates explain the variance within the estimation sample (e.g., through ANOVA or fixed effect models), out-of-sample estimates assess the model's ability to predict unseen data. To make results comparable, we report the GP effect as the share of explained out-of-sample variance that is gained by including GP fixed effects in our ML models (see Shimao et al. 2025).

Finally, we evaluate the predictive contribution of VC firm effects with the random forest algorithm and find that their inclusion improves out-of-sample R^2 by 2.6 percentage points. Hence, the GP effect is greater than the VC firm effect (~38% larger) and the difference is statistically significant, indicating that analyses at the firm level (e.g., Fitza et al. 2009, Nanda et al. 2020) may obscure meaningful performance variations arising from individual GPs. Again, however, this difference is smaller than that reported in existing studies (e.g., Ewens and Rhodes-Kropf 2015), which find a five-fold difference between GP and VC firm effects.

This paper makes a number of contributions to the literatures in strategy, entrepreneurship and entrepreneurial finance on the sources of performance persistence in VC. First, we advance current methodological approaches, almost exclusively based on linear regression methods and variance decomposition, to detect the effect of GPs on VC performance persistence. In so doing, we demonstrate that traditional in-sample methods are problematic and that GP effects should be resized. Second, we move beyond the traditional focus on VC firm attributes as a driver of performance (e.g., Harris et al. 2023, Kaplan and Schoar 2005) by assessing whether *individual* factors matter for performance persistence. Our finding that GP effects consistently exceed VC firm effects extends earlier work by Ewens and Rhodes-Kropf (2015) and contributes to the debate on organizational versus human capital determinants of performance persistence in VC (e.g., Zarutskie 2010). Third, we contribute to the broader literature on the effect of managers on firm performance (e.g., Bertrand and Schoar 2003, Demerjian et al. 2012, Fitza 2014) by documenting a significant, albeit economically small effect in the VC context. The VC industry provides a useful setting to examine the presence of managerial effects because GPs operate funds where their decisions can directly affect returns through venture selection, guidance and exit timing. We find a small but statistically significant effect, in contrast with recent evidence from studies on public firms' CEOs, which report virtually no CEO effect when using ML methods (Shimao et al. 2025). Generally, this difference suggests that managerial effects are highly context-dependent, which is in itself a novel finding to the literature.

The paper is organized as follows. Section 2 positions our work within the literature and presents a brief review of the GP and VC firm effects in existing research, also linking it to the debate on CEO effects. Section 3 introduces the ML methodology. Section 4 illustrates the data, and Section 5 presents our main results. Sections 6 and 7 discuss our findings and conclude.

2. Review of Existing Methods and Results

Although research on the effects of GPs in the VC industry is not as developed as, for example, that on the effects of CEOs in public firms, which dates back to, e.g., Hambrick and Mason (1984), a considerable stream has emerged within a literature at the crossroads of strategy, entrepreneurship and finance. In an industry characterized by substantial uncertainty (Amit et al. 1998, Ewens et al. 2018), scholars have examined whether some VC firms systematically outperform their peers over time, and what might be the sources of such performance persistence (e.g., Kaplan and Schoar 2005, Korteweg and Sorensen 2017), also going back to one of strategy's core questions concerning how to sustain competitive advantage (Barney 1991, Porter 1985). In what follows, we review the research on performance persistence in the VC industry paying particular attention to the roles of GPs and VC firms.[6] This allows us to situate our work within the existing literature and motivate our contributions.

Initial studies in this domain were typically conducted at the level of the VC firm because detailed data on the affiliations of people working in the VC industry were hardly available at the time (Da Rin et al. 2013). Kaplan and Schoar (2005) investigate performance persistence in the VC industry between 1980 and 2001 using Venture Economics and show that the inclusion of VC firm FE improves explanatory power (i.e., in-sample R^2). The authors argue that this evidence is suggestive of the presence of unobservable, time-invariant characteristics of the VC firm that might systematically influence fund performance, which they speculatively trace back to GPs' human capital.[7] Fitza et al. (2009) use

[6] Given the dynamic nature of the VC industry, where GPs begin raising the subsequent fund while the previous fund is still open (Barber and Yasuda 2017), performance is often studied dynamically. As a consequence, it is not uncommon that the boundary between performance and performance persistence is blurred and the two are investigated within the same research setting and treated as highly related. For instance, several authors look at the relationship between initial and subsequent performance with a focus on both VCs' performance and their persistence (e.g., Cong and Xiao 2022, Kaplan and Schoar 2005, Nanda et al. 2020, Zarutskie 2010). On this note, we think that it may be helpful to recognize a theoretical difference between these two concepts, whereby performance is a static concept while performance persistence is a dynamic concept. In the following, we highlight whether the reviewed articles focus on performance persistence, i.e., implicitly adopting a dynamic view of VC performance, (see Table 2). We show that all but one article investigate performance persistence and, even this outlier (i.e., Fitza et al. 2009) accounts for persistence in its analyses.

[7] Kaplan and Schoar (2005) refer to VC firms using the term GPs, which may be confusing in regard to the theoretical framing used in our paper, in which GP refers to the individual manager that operates a VC fund.

variance decomposition methods to estimate the share of variance in the performance of early-stage, VC-backed ventures that can be attributed to different sources, among which the VC firm. They employ ANOVA on inter-round venture valuation changes using data from VentureXpert spanning 1980 to 2005. They report that the FE at the VC firm level explain around 11% of the total variance in outcomes, providing additional empirical support for the presence of an effect specific to the VC firm.

Among the first to move the analysis to the individual GP level, Ewens and Rhodes-Kropf (2015) study performance persistence leveraging changes in GP-VC firm affiliations over time in order to estimate both VC firm and GP FE using investment-level data over 1987 to 2012. The authors find that the GP effect accounts for 47% of the variance in the persistence of ventures' exit valuations and is up to five times larger than the VC firm effect. They argue that this shows how GPs, and not just the firms that employ them, can influence venture outcomes via repeatable investment skills, providing empirical support to the arguments of Kaplan and Schoar (2005). Our study is close to Ewens and Rhodes-Kropf (2015) and extends that paper in several ways. First, we employ a different performance variable (i.e., fund-level IRR), which provides a more comprehensive and financially grounded measure of VC fund performance compared to investment-level outcomes. Second, while their identification strategy relies primarily on GPs who move across VC firms to estimate GP and VC firm FE, our fund-level design captures variation across multiple funds under the same VC firm but managed by teams of different GPs, allowing identification of the GP effect without depending exclusively on movers.[8] Third, we adopt ML methods mitigating the risk of overfitting inherent in FE analyses and, as a result, our estimated GP effect is smaller but likely represents a more conservative and reliable estimate of the real explanatory power of GPs. Overall, we analyze a larger and more recent dataset and use measures and methods that enhance empirical robustness and improve predictive validity. In doing so, we confirm that GPs contribute to performance persistence (more so than VC firms), though the magnitude of their influence appears to be much smaller than existing estimates.

Several other papers offer theoretical and empirical rationales for why performance persistence might arise even when GPs lack skill, which, as noted earlier, does not necessarily imply that GPs do

[8] We also conduct analyses on a sub-sample of data that only includes mobile GPs, which enables us to leverage the strictest separation between GPs and VC firms. These analyses confirm our main results.

not matter. Hochberg et al. (2013) present a model of informational hold-up, where VCs' informational advantages deter syndicate partners from defecting, reinforcing network stability and performance persistence. Empirically, they show that GPs with past success attract more syndicate partners and can raise larger follow-on funds, consistent with a view that (past success of) GPs matters to explain performance persistence. Relatedly, Nanda et al. (2020) show that early IPO success increases access to higher-quality syndicates and better investment opportunities. These patterns can generate persistent performance without necessarily relying on GP skill as the main explanatory variable but attributing a central role to GPs nonetheless. Cong and Xiao (2022) propose a theory in which persistence emerges from dynamic LP-GP-startup contracting and matching processes. In their framework, even when GPs are ex-ante identical, random luck and endogenous mechanisms (e.g., continuation contracts, tolerance for short-term failure) generate GP heterogeneity leading to performance persistence. While these articles challenge the interpretation of persistence as direct evidence of GPs' skills, they argue, however, that GPs have a central role in generating persistent returns through path-dependent dynamics.

On a general note, the level of analysis of performance persistence remains an unresolved issue in the VC literature. While many articles investigate performance persistence at the level of the VC firm (e.g., Gompers et al. 2009, Harris et al. 2023, Hochberg et al. 2007, Kaplan and Schoar 2005, Nanda et al. 2020), there is evidence that the individual characteristics of GPs are correlated with performance (e.g., Bottazzi et al. 2008, Dimov and Shepherd 2005, Zarutskie 2010). Our findings align with this second line of inquiry by showing that GPs consistently explain more out-of-sample variance in funds' performance persistence compared to VC firms, complementing the theoretical distinction between organizational and human capital in VC (Ewens and Rhodes-Kropf 2015).

Finally, Shimao et al. (2025) emphasize the importance of validating managerial effects using ML methods. The authors criticize traditional FE approaches used to estimate manager effects in the CEO literature (e.g., Bertrand and Schoar 2003, Crossland and Hambrick 2007, Fitza 2017, Mackey 2008) for their limited external validity. Using ML techniques, Shimao et al. (2025) demonstrate that many estimated CEO effects are inaccurate (a finding in line with, e.g., Jarosiewicz and Ross 2023). In prior work on the CEO effect, Fitza (2014) highlights the methodological risks of standard variance decomposition analyses. He observes that the share of performance variance attributed to CEOs may be inflated by the limited number of CEO-specific observations, which leaves ample room for random variation to generate substantial apparent manager effects. A similar critique is implicit in Korteweg and Sorensen (2017), which investigate performance persistence in private equity (PE) and find that while PE firms consistently exhibit different levels of performance persistence, realized returns are extremely noisy and hard to predict ex ante. These considerations suggest that the literature on GP effects may benefit from employing ML (i.e., predictive) methods vis-à-vis traditional methods.

Insert Table 2 about here

In Table 2, we summarize the papers that investigate the role of GPs or VC firms in explaining performance persistence. For each paper, we report whether it estimates a GP or VC firm effect, whether it studies performance persistence, the adopted performance measure(s), the unit of analysis, the methodology used, the data source(s), the sample size and the time span. As shown, many existing works focus on firm-level variation, typically using linear models such as FE or ANOVA to evaluate in-sample variation. A smaller number of studies moves toward identifying individual GP effects but still relies on in-sample inference. These approaches may have limited external validity, as emphasized by critiques in both the VC and broader manager effects literatures. Our study departs from existing ones by leveraging recent, detailed data on VC firms, funds and GPs, in addition to using ML algorithms to test whether the inclusion of GP-level identification improves predictive accuracy, providing a more rigorous empirical foundation for the GP effect in the VC industry.

3. Methods

As previously mentioned, in-sample linear methods (e.g., Fitza et al. 2009, Korteweg and Sorensen 2017) suffer from overfitting, and when the estimated parameters are applied to new data, they fail to provide accurate predictions, particularly with large datasets or in the presence of high collinearity (Hastie et al. 2009, Bishop 2006). To address this issue and isolate the persistent effect (if any) of GPs on performance, we operationalize this problem as an out-of-sample prediction problem and employ a set of ML algorithms (following Shimao et al. 2025). To measure the GP and VC firm effects, we consider three alternative model specifications: (1) a baseline model with fund-specific controls (e.g., vintage year, LPs' capital commitment) and equity market returns with four lags; (2) a GP-enhanced model that adds GP FE to the baseline; and (3) a VC-enhanced model that adds VC firm FE to the baseline. To ensure sufficient heterogeneity, reduce data sparsity and uncover potential effects, we require GPs and VC firms to have run at least 2 funds and present 40 quarterly observations. For our main analyses, we select the best-performing ML algorithm based on its out-of-sample performance for the baseline model specification.

Importantly, our results do not depend on model selection since we employ a large set of ML algorithms: regularized linear algorithms (i.e., ridge regression, least absolute shrinkage and selection operator (LASSO) and elastic net combining L1/L2 penalties), tree-based ensemble methods (i.e., random forest, extreme gradient boosting (XGBoost) and light gradient boosting machine (LightGBM)) and kernel methods (i.e., support vector machines (SVM) with radial basis function kernels) (Breiman 2001, Chen and Guestrin 2016, Friedman 2001). If different ML approaches return similar results, it means that whatever effect is detected, it is unlikely to be an artifact of a specific modeling choice (Hastie et al. 2009, Shimao et al. 2025).

We implement a forward-looking time-series cross-validation (Hyndman and Athanasopoulos 2018) based on an expanding training window. The first 20 quarters constitute the initial training set, while the 21st quarter serves as the first test set. In each subsequent iteration, the training set expands by one quarter (thus including quarters 1-21, then 1-22, etc.), while the test set remains the quarter immediately following the training period. In this process, we use 80 folds in total, representing 20 years of out-of-sample predictions. This choice balances two concerns. First, we require a sufficiently large initial training set to reliably estimate GP and VC firm effects, since a training window that is too narrow would result in a high-dimensional trap where the number of parameters is larger than the number of observations. Second, we also need sufficient test periods to ensure that results do not depend on specific market conditions or time periods. Time-series cross-validation appropriately accounts for temporal dependence and prevents the look-ahead bias that could artificially inflate performance estimates in our setting (Bergmeir and Benítez 2012).

Following Shimao et al. (2025), the GP effect is measured in terms of percentage improvement in out-of-sample R^2 of the GP FE model compared to the baseline:

$$GP_{Effect} = \frac{R^2_{\{GP-enhanced\}} - R^2_{\{Baseline\}}}{R^2_{\{Baseline\}}} \times 100\%, \tag{1}$$

and, similarly, the VC effect is the percentage improvement in out-of-sample R^2 of the VC firm FE model compared to the baseline:

$$VC\ firm_{Effect} = \frac{R^2_{\{VC\,firm-enhanced\}} - R^2_{\{Baseline\}}}{R^2_{\{Baseline\}}} \times 100\%. \tag{2}$$

We test the statistical significance of the magnitude of improvement over the baseline model with paired t-tests on the R^2 series from the cross-validation procedure (Shimao et al. 2025). To account for temporal dependence, heteroskedasticity (MacKinnon 2006) and avoid normality assumptions, we also implement a wild bootstrap version of paired t-tests (Davidson and Flachaire 2008). Finally, since we evaluate effects across multiple ML algorithms, as a robustness test, we apply Benjamini and Hochberg (1995) corrections to control for false positives.

4. Data

We construct our sample using PitchBook's universe of VC funds. PitchBook is a leading industry database commonly used in the entrepreneurial finance and strategy literatures (e.g., Amore et al. 2023, Hong and Mella-Barral 2024, Vollon 2025) from which we gather quarterly fund performance metrics, fund characteristics, GP-VC fund matches and VC firm-VC fund matches necessary to test our research question at the fund-quarter unit of analysis. We initially track 13,091 unique GPs across 14,116 funds. We then obtain a sub-sample of 811 GPs and 217 VC firms based on our two criteria (i.e., they must have managed at least 2 funds and be observed for at least 40 quarters). The 40 quarters threshold matches a standard VC fund's lifespan (Metrick and Yasuda 2010, Phalippou 2010), without specifically requiring a fund to be classified as liquidated. From an econometric point of view, focusing on GPs with sufficient track-record prevents the high-dimensional trap and collinearity issues that may arise with sparse data (Wooldridge 2010). We acknowledge that these two conditions introduce potential selection bias toward more established and successful GPs as we may exclude those who exit the industry early, potentially due to poor performance. This may lead us to overestimate the GP effect in cases where unsuccessful GPs manage fewer funds. However, this restriction is necessary for identification and a prerequisite for distinguishing GP effects from fund-specific effects.[9]

Our final sample comprises 29,021 VC fund-quarter observations spanning 1997-2022 for 811 GPs who manage 722 funds in 217 VC firms. The large cross-sectional sample size and the temporal range, covering multiple market cycles, provide a robust foundation for ML analyses (Phalippou and Gottschalg 2009). The sample is globally representative but skewed towards the United States (i.e., 92.6% of the observations) (Lerner and Nanda 2020). An important caveat in this context involves attributing performance when GPs move between VC firms or manage funds at multiple firms. As displayed in Table 3, 35% of GPs manage funds at more than one VC firm over the sample period, with these mobile GPs accounting for 42.6% of all fund-GP observations. Mobile GPs demonstrate higher experience, managing an average of 6.4 funds compared to 4.6 funds for single-firm GPs. For the scope of our GP effect analyses, we treat each GP-fund pair separately, meaning that a GP's FE follows the individual across quarters and funds. This means that performance patterns are tracked at the individual GP level regardless of their VC firm affiliation at any given quarter. Conversely, for the VC firm effect analysis, FE relate to the VC firm regardless of who manages it. Each fund contributes an average of 40.2 quarterly observations, consistently with the average expected fund lifecycle of 10 years.

Insert Table 3 about here

Our performance variable is the quarterly IRR, which is the industry-standard performance metric in VC (Harris et al. 2014, Kaplan and Schoar 2005). We interpolate missing intermediate quarters when gaps in IRR are five periods or fewer. Unlike other measures (e.g., the number of exits or the exit rate of a fund), the IRR captures the actual financial performance that industry actors like LPs evaluate (Harris et al. 2023). Our data exhibits high variance, typical of VC returns (Cochrane 2005), with the IRR averaging 6.51% but ranging from as low as -99.9% to as high as 330%. Across our 811 GPs, the

[9] A more thorough discussion of this issue is contained in Section 6.1.

mean IRR per GP averages 8.4%, with the median GP achieving 5.5%, while bottom-quartile GPs average -1.7% and top-quartile GPs average 15.6%. VC firms show even greater dispersion whereby the mean IRR per firm averages 4.9%, with the median firm at 2.7%, bottom-quartile firms at -5.3% and top-quartile firms at 12.4%. Notably, GP-level IRRs exceed VC firm-level IRRs across the distribution, which empirically motivates our investigation. The main features of interest are GP (VC firm) FE that are dummies identifying matches between GPs (VC firms), time periods and VC funds. We encode GP (VC firm) FE through sparse matrices (i.e., 29,021 × 811 for GPs, and 29,021 × 217 for VC firms), where each column represents a unique GP (VC firm). This approach captures all available GPs' and VC firms' associations with fund-quarters without imposing any restriction on variable selection but, at the same time, exposes results to overfitting when relying on standard in-sample linear models (Tibshirani 1996, Zou and Hastie 2005), motivating the use of ML methods.

We include several features to account for heterogeneity in observable fund characteristics. We quantify the financial endowment of a fund using *Fund size*, which measures the commitment of LPs in millions of US dollars. We also include *GP team size*, that is, the number of GPs involved in managing the fund, *VC size*, which measures the number of VC firms partnering on the fund and *LP size*, that measures the number of LPs who committed capital to the fund. We control for fund lifecycle and maturity using two variables: *Fund age*, measured in quarters from inception, and *Lifecycle stage*, a set of indicator variables that highlight the operational stage of the VC fund (i.e., early stage, growth stage and harvest stage). We account for the location of the fund at the country level using FE. We include controls for public market conditions through returns and lagged returns of the NASDAQ Composite Index[10], using up to four quarterly lags to ensure that we comprehensively capture the influence of market dynamics on fund performance since public market conditions may affect VC returns with varying delays (Gompers et al. 2020). We also include FE for year, quarter and vintage year (i.e., the year in which the fund starts making investments) to control for timing effects.

----

[10] The NASDAQ Composite Index is widely used as a benchmark for VC returns in both academic literature and industry practice. As an example, Cambridge Associates systematically compares their US VC Index against the NASDAQ Composite Index in all their benchmark reports (see https://www.cambridgeassociates.com/en-eu/private-investment-benchmarks/, consulted online on November 29, 2025).

Insert Table 4 about here

In Table 4, we provide an overview of the summary statistics of our features. Most GPs manage relatively few funds during their careers, with nearly half (namely, 48.5%) managing exactly two funds and the majority (namely, over 70%) managing four or fewer funds. The skewed and right-tailed distribution of the measure of GPs' experience provides a strong motivation for the identification of a GP effect since the most experienced GPs (i.e., those managing 10 or more funds) represent less than 2%. Also, this ensures that our results are not driven by a small subset of star GPs (i.e., the highly experienced ones). This distribution aligns with observed career patterns in the VC industry, where fundraising cycles typically span three to five years and successful GPs may manage three to six funds over their career (Kaplan and Schoar 2005). The *Lifecycle stage* distribution shows that our sample captures funds across all investment phases, from early-stage funds (i.e., 9.1%) to growth-stage funds (i.e., 30.8%) and harvest-stage funds (i.e., 60.1%). Our sample encompasses the full range of funds, from initial capital deployment through portfolio maturation and exit activities (Cumming and Johan 2013).

5. Findings

In this section, we present our findings on the persistent effects of GP and VC firms on VC performance. The Appendix includes the importance analysis and the robustness analysis across all ML algorithms.

5.1 Issues with Linear Models

We start by highlighting the large extent to which linear models overfit in our setting. As a preliminary test, we run an OLS-equivalent ridge regression with minimal penalization (λ=0.001), which implies negligible shrinkage on coefficients. With this estimation method, the baseline model achieves 38.5% in-sample R^2 and only 4.3% of out-of-sample R^2. Similarly, the GP-enhanced model achieves an in-sample R^2 of 74.2% and an out-of-sample R^2 of 43.5%. Finally, the VC-enhanced model yields 56.1%

in-sample R^2 and only 33% out-of-sample R^2. Based on these results, we can conclude that when in-sample coefficient estimates are applied out-of-sample, linear models overfit and fail to provide accurate predictions, thereby having limited external validity.[11]

5.2 Model Selection

In Table 5, we report out-of-sample R^2 values for all ML algorithms (see also Figure 1). We select random forest as our best-performing algorithm based on the out-of-sample R^2 in the baseline model specification. Random forest outperforms other approaches like SVM, XGBoost or LightGBM. As anticipated, linear models demonstrate very poor R^2 in the baseline specification, with ridge achieving a negative out-of-sample R^2, similarly to LASSO and elastic net. The higher performance of non-linear algorithms seems to imply that the relationship between fund features and IRR involves non-linear interactions that linear shrinkage models fail to capture (Breiman 2001). This difference also justifies the need for sophisticated ML approaches to accurately predict VC fund performance, even if only a single quarter ahead.

Insert Table 5 and Figure 1 about here

5.3 GP and VC Firm Effects

In this section, we present evidence to address our research question on the persistent effect of GPs (and VC firms) on VC fund performance (see Table 6). In what follows, we focus on the random forest

[11] Anticipating some results (see Sections 5.2 and 5.3), we also observe large gaps between linear and non-linear models in all specifications. Linear algorithms achieves very low R^2 values in the baseline model specification due to a poor signal-to-noise ratio. Indeed, regularization shrinks coefficients toward zero because it cannot find a stable linear relationship between fund features and IRR. As a result, the negative R^2 for the ridge algorithm indicates that it is performing worse than just predicting the mean. On the contrary, GP and VC firm FE model specifications consist of hundreds of sparse binary features and regularized linear models excel at this type of high-dimensional selection problem. Regularized models can effectively pick GPs and VC firms that have strong predictive power while shrinking out the rest. Non-linear algorithms instead perform well even in the baseline specification, suggesting that the relationship between fund features and IRR is highly non-linear. These findings confirm that linear algorithms, even with shrinkage, are not the best option for detecting GP or VC firm effects. Without parameter tuning, they strongly overfit, while with parameter tuning, they fail to capture non-linearities and therefore overestimate effects. These results suggest that previous studies relying on linear methods may have overestimated the size of GP effects relative to fund features.

algorithm only, since it emerged as the best-performing. The GP effect calculated by comparing the GP-enhanced specification with the baseline specification consists of a 3.6% improvement in out-of-sample R^2 (computed as 77.8% – 74.2%). The increase is statistically significant based on both our inferential approaches (i.e., standard and wild bootstrap t-tests). According to this result, we confirm that GPs have a persistent effect on VC funds' performance given that the inclusion of GP FE contributes to predicting a fund's IRR more accurately. Perhaps surprisingly, but in line with recent evidence (Shimao et al. 2025), the size of this effect is small and significantly lower than previously documented.

Insert Table 6 about here

To contextualize the economic magnitude of this effect, consider an LP selecting among VC funds. In our sample, quarterly IRRs have a standard deviation of 26.7%, hence the baseline model's R^2 of 74.2% translates to an RMSE of 13.56%. The GP-enhanced specification improves the R^2 to 77.8%, reducing the RMSE to 12.58%. This represents a reduction in prediction error of ~1%, or a 7.2% relative improvement in forecasting accuracy. Now, for an LP making capital allocation decisions, this precision could translate to better portfolio construction and risk management. For example, choosing a fund with a 3% quarterly IRR versus one with 2.9% quarterly IRR (i.e., just a 10-basis point delta) results in very different returns over a typical 10-year fund life, i.e., $326.2 million versus $313.8 million.

Moreover, GP effects consistently exceed VC firm effects. The VC firm effect calculated comparing the VC-enhanced specification with the baseline specification shows a 2.6% improvement in out-of-sample R^2 (calculated as 76.8% – 74.2%). The two-tailed paired t-test performed on the out-of-sample R^2 of the VC firm-enhanced specification shows statistical significance both when compared to the baseline specification and to the GP-enhanced specification. Hence, both the GP and the VC firm effects improve predictive power over the baseline, but the GP effect is significantly greater than the VC effect. This finding has relevant implications for the VC literature, which has traditionally focused on firm-level analyses (Gompers et al. 2020, Kaplan and Schoar 2005) and supports existing in-sample evidence (Ewens and Rhodes-Kropf 2015).

Insert Figure 2 about here

In Figure 2, we show the wild bootstrap null distributions of the GP and VC firm effects, employed to account for temporal dependence in the cross-validated R^2 series. Under the null hypothesis of no effect, the bootstrap procedure generates distributions centered at zero by randomly flipping the signs of the observed differences across test periods. The GP effect of 3.6% and the VC firm effect of 2.6% fall well outside the mass of these null distributions. In both cases, fewer than 5% of the bootstrap resamples exceed the observed effect in absolute value, which confirms statistical significance at conventional levels.

The stronger GP effect relative to the VC firm effect is particularly meaningful in light of the high GP mobility in our sample. With 46.2% of fund-GP observations involving mobile GPs who work across multiple firms, the portable nature of GP-specific idiosyncrasies appears to dominate firm-specific advantages like resources or reputational capital. In order to address the role and potential bias of mobile GPs, we re-run the random forest algorithm on the subset of mobile GPs. While still improving over the baseline, the effect of mobile GPs is not statistically different from that of the VC firm and the overall GP effect (i.e., our main result, including mobile and non-mobile GPs). In other words, we have evidence showing that the main GP effect is not driven by a small subset of the same star GPs contended across VC firms.

6. Discussion

In this paper, we examined whether GPs contribute to explaining VC fund performance over time, what we call the persistent effect of GPs. While the link between GPs and performance persistence is often taken for granted in the VC literature, whether through differences in skill (Kaplan and Schoar 2005) or as a result of luck (Cong and Xiao 2022, Nanda et al. 2020), there has been limited empirical work testing this relationship. We addressed this gap using ML and employing a large dataset of 29,021 quarterly fund-level observations from PitchBook, covering 722 funds managed by 811 GPs between 1997 and 2022.

Our findings indicate that GPs do matter albeit to a much smaller extent than previously thought. In nearly all the ML algorithms that we tested, the addition of GP FE leads to better out-of-sample predictions when compared to the baseline specification, meaning that GPs have predictive power or, in other words, that they exhibit performance persistence. For our best-performing algorithm (i.e., random forest), we showed that the out-of-sample R^2 increases from 74.2% to 77.8% when GP FE are added. This increase is statistically significant but far below the effects identified in extant literature (e.g., Ewens and Rhodes-Kropf 2015). We also compared the GP effect against the VC firm effect. In every ML algorithm, including our best-performing model, adding VC firm FE significantly improves prediction, but to a smaller extent compared to GP FE. This result suggests that while VC firms matter as well, GPs account for a larger share of the persistence in fund outcomes.

Central to our approach is the use of out-of-sample prediction. Previous studies, including those in the VC and CEO literatures, with the exception of Shimao et al. (2025), have used linear models to estimate manager effects. While informative, these methods rely on in-sample variance only, which limits their external validity and has led to conflicting findings. Linear models can overstate the importance of individual managers by capturing noise rather than meaningful signal (e.g., Jarosiewicz and Ross 2023), something that can be mitigated with the use of ML methods. Our analysis revealed that the choice between linear and non-linear methods, even with ML, is not only technical but also fundamental for external validity. By using non-linear ML algorithms, which can flexibly handle high-dimensional and sparse data, we identified a significant GP effect, which is smaller in size compared to extant research but likely more accurate.

Our results contribute to three strands of literature. First, we add to the works that study VC performance persistence by exploring the underlying assumption that GPs play a central role. Prior studies have documented persistence in VC returns (e.g., Harris et al. 2023, Kaplan and Schoar 2005) and have suggested that this can be traced back to GPs (Ewens and Rhodes-Kropf 2015). We confirmed this by showing that our GP-enhanced model specification consistently outperforms other model specifications, yet the GP effect needs to be resized compared to prior literature. Second, our findings inform the debate on the relative importance of human versus organizational capital in VC firms (Ewens and Rhodes-Kropf 2015). While earlier, but also more recent studies (Nanda et al. 2020), focus on VC firms, we showed the importance of focusing on individual managers, who are characterized by a performance premium consistently larger than that of their employing organizations. On this note, we also showed that this effect is not determined by GPs' mobility across VC firms. Third, we contribute to the broader manager effects literature, dating back to the influential study by Bertrand and Schoar (2003), by documenting an effect of individual managers on multiple funds in the VC industry. We believe that our setting differs from commonly studied public firms, where it may be difficult to isolate the influence of managers (e.g., CEOs) and, as a result, estimate their effect (see Shimao et al. 2025). GPs operate in smaller, more autonomous funds and directly make capital allocation decisions (Tyebjee and Bruno 1984), providing us with performance outcomes that are easier to attribute to their management.

We also believe that our findings have practical implications for professionals in the VC industry. The higher predictive power of GP-level information suggests that LPs and VC firms could benefit from systematically tracking individual GPs. Our results indicate that considerations based solely on the VC firm's historical performance may be incomplete if key GPs have left. For instance, given that GP effects exceed VC firm effects, LPs might consider the stability and composition of GP teams as carefully as they evaluate firm-level track records.

6.1 Limitations and Future Research

Our sample selection criteria warrant additional discussion. By requiring GPs to have managed at least two funds with 40 quarters of observation each, we restrict our focus to more established managers who remain in the industry long enough to raise at least two funds. This requirement, while essential for identifying GP effects separate from fund-specific variation, means that our results apply to a subset of GPs rather than the entire population. The persistent effect of GPs that we identify may differ for early-career GPs or those who exit the industry after managing a single fund. This could lead to overestimation of the GP effect if unsuccessful GPs have systematically lower predictive power. We acknowledge that GPs excluded from our sample may still possess predictive characteristics and that our restriction is driven by identification requirements rather than assumptions about these specific GPs. Future research could explore whether and which GP characteristics observable in first funds predict subsequent performance (see Zarutskie 2010 for an early attempt in this direction).

7. Conclusion

In this paper, we show that GPs play a significant role in explaining variation in VC performance persistence over time, yet their influence is significantly smaller than previously suggested. In particular, our results indicate that the persistent effect of GPs is greater than that of the VC firms employing them (similar to Ewens and Rhodes-Kropf 2015) but also that a large share of fund performance depends on broader structural and market factors. By providing a robust estimate of the GP effect via out-of-sample prediction and ML, we refine existing evidence on performance persistence in VC and provide reliable magnitudes of the influence of GPs on fund outcomes. Overall, our findings suggest that managerial effects are present and systematic in the VC industry, differently from recent evidence on the CEOs of public firms (Shimao et al. 2025), but that their magnitude was overestimated in extant research.

References

Amit R, Brande J, Zott C (1998). Why do Venture Capital firms exist? Theory and Canadian evidence. *Journal of Business Venturing, 13(6): 441-466*.

Amore MD, Conti A, Pelucco V (2023). Micro venture capital. *Strategic Entrepreneurship Journal, 17(4): 886-924*.

Barber BB, Yasuda A (2017). Interim fund performance and fundraising in private equity. *Journal of Financial Economics, 124(1): 172-194*.

Barney J (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17(1): 99-120*.

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, 57(1): 289-300*.

Bennedsen M, Pérez-González F, Wolfenzon D (2020). Do CEOs matter? Evidence from hospitalization events. *Journal of Finance, 75(4): 1877-1911*.

Bergmeir C, Benítez JM (2012). On the use of CV for time series predictor evaluation. *Information Sciences, 191: 192-213*.

Bertrand M, Schoar A (2003). Managing with style: The effect of managers on firm policies. *Quarterly Journal of Economics, 118(4): 1169-1208*.

Bishop CM (2006). Pattern recognition and machine learning. *Springer, New York*.

Black, J. A., & Boal, K. B. (1994). Strategic resources: Traits, configurations and paths to sustainable competitive advantage. *Strategic Management Journal*, *15*(S2), 131-148.

Breiman L (2001). Random forests. *Machine Learning, 45: 5-32*.

Chen T, Guestrin C (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: 785-794*.

Cochrane JH (2005). The risk and return of venture capital. *Journal of Financial Economics, 75(1): 3-52*.

Cong, LW, Xiao Y (2022). Persistent blessings of luck: Theory and an application to venture capital. *Review of Financial Studies, 35(3): 1183-1221*.

Crossland C, Hambrick DC (2007). How national systems differ in their constraints on corporate executives: A study of CEO effects in three countries. *Strategic Management Journal, 28(8): 767-789*.

Cumming D, Johan S (2008). Information asymmetries, agency costs and venture capital exit outcomes. *Venture Capital, 10(3): 197-231*.

Cumming D, Johan S (2013). *Venture capital and private equity contracting: An international perspective.* Academic Press, Burlington.

Da Rin M, Hellmann T, Puri M (2013). A survey of venture capital research. *Handbook of the Economics of Finance,* Elsevier, Amsterdam.

Davidson R, Flachaire E (2008). The wild bootstrap, tamed at last. *Journal of Econometrics, 146(1): 162-169*.

Demerjian P, Lev B, McVay S (2012). Quantifying managerial ability: A new measure and validity tests. *Management Science, 58(7): 1229-1248*.

Dimov DP, Shepherd DA (2005). Human capital theory and venture capital firms: Exploring ‘‘home runs’’ and ‘‘strike outs’’. *Journal of Business Venturing, 20(1): 1-21*.

Ewens M, Nanda R, Rhodes-Kropf M (2018). Cost of experimentation and the evolution of venture capital. *Journal of Financial Economics, 128(3): 422-442*.

Ewens M, Rhodes-Kropf M (2015). Is a VC partnership greater than the sum of its partners? *Journal of Finance, 70(3): 1081-1113*.

Fee CE, Hadlock CJ, Pierce JR (2013). Managers with and without style: Evidence using exogenous variation. *Review of Financial Studies*, *26(3): 567-601*.

Fitza MA (2014). The use of variance decomposition in the investigation of CEO effects: How large must the CEO effect be to rule out chance? *Strategic Management Journal, 35(12): 1839-1852*.

Fitza MA (2017). How much do CEOs really matter? Reaffirming that the CEO effect is mostly due to chance. *Strategic Management Journal, 38(3): 802-811*.

Fitza MA, Matusik SF, Mosakowski E (2009). Do VCs matter? The importance of owners on performance variance in start-up firms. *Strategic Management Journal, 30(4): 387-404*.

Friedman JH (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29(5): 1189-1232*.

Gompers P, Gornall W, Kaplan SN, Strebulaev IA (2020). How do venture capitalists make decisions? *Journal of Financial Economics, 135(1): 169-190*.

Gompers P, Kovner A, Lerner J (2009). Specialization and success: Evidence from venture capital. *Journal of Economics & Management Strategy, 18(3): 817-844*.

Gompers P, Lerner J (2001). The venture capital revolution. *Journal of Economic Perspectives, 15(2): 145-168*.

Gorman M, Sahlman WA (1989). What do venture capitalists do? *Journal of Business Venturing, 4(4): 231-248*.

Hambrick DC, Mason PA (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review, 9(2): 193-206*.

Harris RS, Jenkinson T, Kaplan SN (2014). Private equity performance: What do we know? *Journal of Finance, 69(5): 1851-1882*.

Harris RS, Jenkinson T, Kaplan SN, Stucke R (2023). Has persistence persisted in private equity? Evidence from buyout and venture capital funds. *Journal of Corporate Finance, 81: 1-16*.

Hastie T, Tibshirani R, Friedman J (2009). The elements of statistical learning: Data mining, inference, and prediction. *Springer, New York*.

Hochberg YV, Ljungqvist A, Lu Y (2007). Whom you know matters: Venture capital networks and investment performance. *Journal of Finance, 62(1): 251-302*.

Hochberg YV, Ljungqvist A, Vissing-Jørgensen A (2013). Informational holdup and performance persistence in venture capital. *Review of Financial Studies, 27(1): 102-152*.

Hong S, Mella-Barral P (2024). Informational holdup by venture capital syndicates. *Journal of Financial & Quantitative Analysis, 59(3): 1362-1400*.

Hyndman RJ, Athanasopoulos G (2018). Forecasting: Principles and practice. *Springer, New York*.

Jarosiewicz VE, Ross DG (2023). Revisiting managerial "style": The replicability and falsifiability of manager fixed effects for firm policies. *Strategic Management Journal*, *44*(3): 858-886.

Kaplan SN, Schoar A (2005). Private equity performance: Returns, persistence, and capital flows. *Journal of Finance, 60(4): 1791.1823*.

Korteweg A, Nagel S (2016). Risk-adjusting the returns to Venture Capital. *Journal of Finance, 71(3): 1437-1470*.

Korteweg A, Sorensen M (2017). Skill and luck in private equity performance. *Journal of Financial Economics, 124(3): 535-562*.

Lerner J, Nanda R (2020). Venture capital's role in financing innovation: What we know and how much we still need to learn. *Journal of Economic Perspectives, 34(3): 237-261*.

Lerner J, Schoar A (2004). The illiquidity puzzle: Theory and evidence from private equity. *Journal of Financial Economics, 72(1): 3-40*.

Lieberson S, O'Connor JF (1972). Leadership and organizational performance: A study of large corporations. *American Sociological Review, 37(2): 117-130*.

Mackey A (2008). The effect of CEOs on firm performance. *Strategic Management Journal, 29(12): 1357-1367*.

MacKinnon JG (2006). Bootstrap methods in econometrics. *Economic Record, 82(s1): S2-S18*.

Metrick A, Yasuda A (2010). The economics of private equity funds. *Review of Financial Studies, 23(6): 2303-2341*.

Nanda R, Samila S, Sorensen O (2020). The persistent effect of initial success: Evidence from venture capital. *Journal of Financial Economics, 137(1): 231-248*.

Phalippou L (2010). Venture capital funds: Flow-performance relationship and performance persistence. *Journal of Banking & Finance, 34(3): 568-577*.

Phalippou L, Gottschalg O (2009). The performance of private equity funds. *Review of Financial Studies, 22(4): 1747-1776*.

Porter ME (1985). Competitive advantage: Creating and sustaining superior performance. *New York: Free Press*.

Shimao H, Kim SJ, Khern-Am-Nuai W, Cohen MC (2025). Revisiting the CEO effect through a machine learning lens. *Management Science, 76(6): 1-13*.

Tibshirani R (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B, 58(1): 267-288*.

Tyebjee TT, Bruno AV (1984). A model of venture capitalist investment activity. *Management Science, 39(9): 1051-1066*.

Vollon LJ (2025). Venture capital and private M&A contracting. *Management Science, forthcoming.*

Wasserman N, Nohria N, Anand B (2001). When does leadership matter? The contingent opportunities view of CEO leadership. *HBS Working Paper No. 02-04.*

Wooldridge JM (2010). *Econometric analysis of cross section and panel data.* Second Edition, MIT Press, Cambridge, MA.

Zarutskie R (2010). The role of top management team human capital in venture capital markets: Evidence from first-time funds. *Journal of Business Venturing, 25(1): 155-172*.

Zou H, Hastie T (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B, 67(2): 301-320*.

Tables and Figures

Table 1. Overfitting of OLS-like Models in In-sample vs. Out-of-sample R^2

Model specification	OLS (in-sample R^2)	OLS (out-of-sample R^2)	Overfitting (%)
Baseline	0.384	0.042	0.342
GP-enhanced	0.741	0.435	0.306
VC Firm-enhanced	0.561	0.330	0.231

In this table, we report the R^2 performance of an OLS-like regression (i.e., a ridge regression with minimal penalization, that approximates OLS behavior while handling the computational singularity) on in-sample vs. out-of-sample data for the three model specifications. It can be noted that the OLS-like model is subject to overfitting as it consistently predicts larger R^2 for in-sample compared to out-of-sample data.

Table 2. Overview of Existing Methods

Paper	**GP / VC firm effect**	**Performance persistence**	**Performance measure**	**Unit of analysis**	**Methodology**	**Data source**	**Sample size**	**Time span**
Ewens and Rhodes Kropf (2015)	Both	Yes	Successful exit (IPO / M&A)	VC funding rounds	Two-way fixed effects (AKM)	VentureSource	27,209 funding rounds 16,897 start-ups 5,324 GPs 3,777 VC firms	1987 - 2012
Fitza et al. (2009)	VC firm	No	Inter-round valuation	VC funding rounds	ANOVA	VentureXpert	6,490 funding rounds 3,756 start-ups 1,418 VC firms	1980 - 2005
Harris et al. (2023)	VC firm	Yes	Financial performance (IRR / MOIC / PME)	VC fund	Quartile regressions	Burgiss	1,408 VC funds	1984 - 2015
Kaplan and Schoar (2005)	VC firm	Yes	Financial performance (IRR / PME)	VC fund	Fixed effects in linear regression	Venture Economics	577 VC funds	1980 - 2001
Korteweg and Sorensen (2017)	VC firm	Yes	Financial performance (IRR / TVPI)	VC fund	Hierarchical linear model (generalized ANOVA)	Preqin	842 VC funds 409 VC firms	1969 - 2001
Nanda et al. (2020)	VC firm	Yes	Successful exit (IPO)	VC funding rounds	Linear regression	VentureXpert	46,013 VC firm-start-up pairs 19,802 start-ups 895 VC firms	1980 - 2016
This paper	Both	Yes	Financial performance (IRR)	VC fund	ML models (e.g., Lasso, regression tree, random forest, etc.)	PitchBook	29,021 VC fund-quarter observations 722 VC funds 811 GPs 217 VC firms	1997 - 2022

Table 3. Information on Mobile GPs

Category	Number of GPs	% of GPs	Avg Funds	% of Observations
Single Firm	527	64.98%	4.69	57.40%
Two Firms	217	26.76%	5.92	29.80%
Three+ Firms	67	8.26%	8.22	12.80%
Mobile GPs (Total)	284	35.02%	6.46	42.60%

In this table, we present information on the mobile GPs (i.e., GPs who move across VC firms) in our data. It can be seen that ~35% of GPs move across VC firms, with ~27% moving across two VC firms and ~8% moving across three or more VC firms. Mobile GPs account for ~42% of our observations.

Table 4. Sample Description and Summary Statistics

Panel A: Sample Composition

Variable	Value
Fund-quarter Observations	29,021
Unique Funds	722
Unique GPs	811
Unique VC Firms	217
Time Period	1997-2022
Average Observations per Fund	40.2
Average Observations per GP	134.17
Average Observations per VC	131.77

Panel B: Fund Performance

Variable	N	Mean	Std Dev	Min	Median	Max
IRR (%)	29,021	6.51	26.73	-99.90	3.48	329.92

Panel C: Fund Characteristics

Variable	N	Mean	Std Dev	Min	Median	Max
Fund Size ($M)	722	407.40	477.02	4	265	3,572.34
GP Team Size	722	6.47	5.62	1	5	54
VC Size	722	1.02	0.14	1	1	2
LP Size	722	12.96	13.87	1	9	159
Fund Age (quarters)	29,021	33.39	20.39	2	30	100

Table 5. Out-of-sample R^2 Performance by each ML Algorithm

ML Algorithm	Baseline Model	GP-enhanced Model	VC Firm-enhanced Model
Random Forest	0.742	0.778	0.768
Support Vector Machine	0.692	0.748	0.726
XGBoost	0.494	0.494	0.497
LightGBM	0.464	0.449	0.459
Elastic Net	0.021	0.533	0.307
LASSO	0.040	0.531	0.300
Ridge	-0.003	0.525	0.316

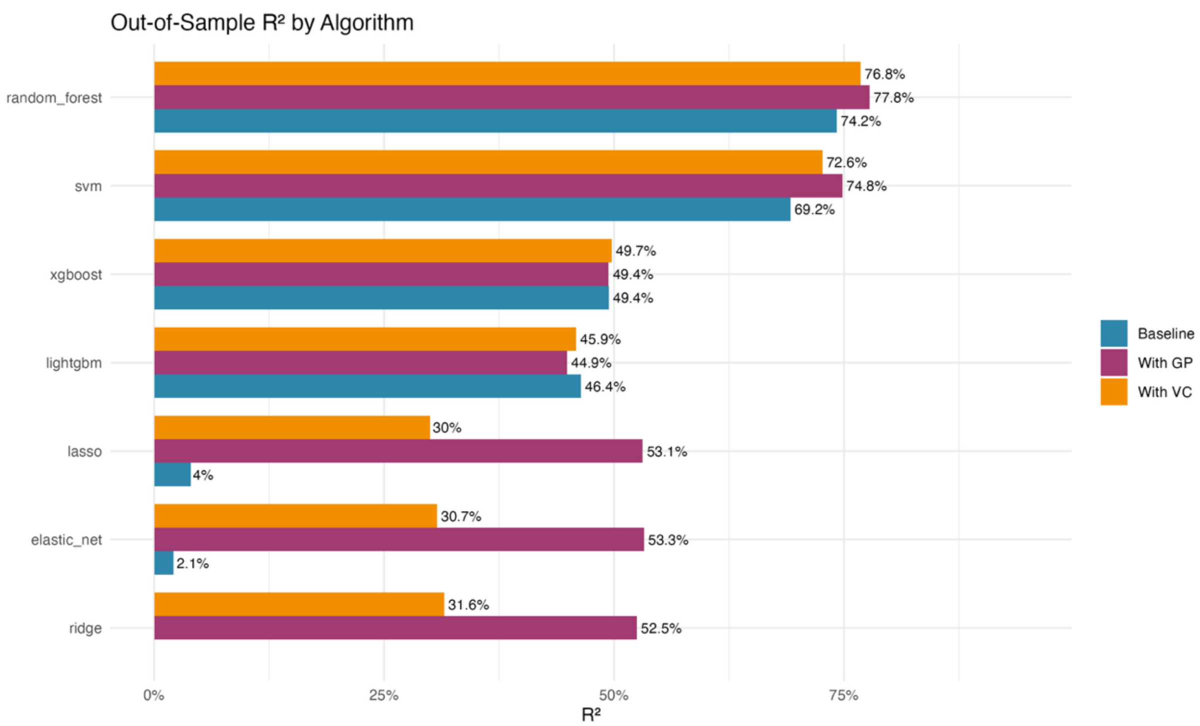
In this table, we present the results for all ML algorithms concerning out-of-sample R^2 predictions of our performance variable (i.e., IRR) for the three model specifications. R^2 values represent out-of-sample predictive accuracy using time series cross-validation with 80 test quarters.

Table 6. Full Results of the Random Forest Model

Model	Out-of-sample R^2	Effect Size	95% CI	t-statistic
Baseline Model	0.742	-	-	-
GP-enhanced Model	0.778	+0.036 (4.80%)	[0.027, 0.043]	8.99^{***}
VC firm-enhanced Model	0.768	+0.026 (3.52%)	[0.019, 0.032]	8.19^{***}

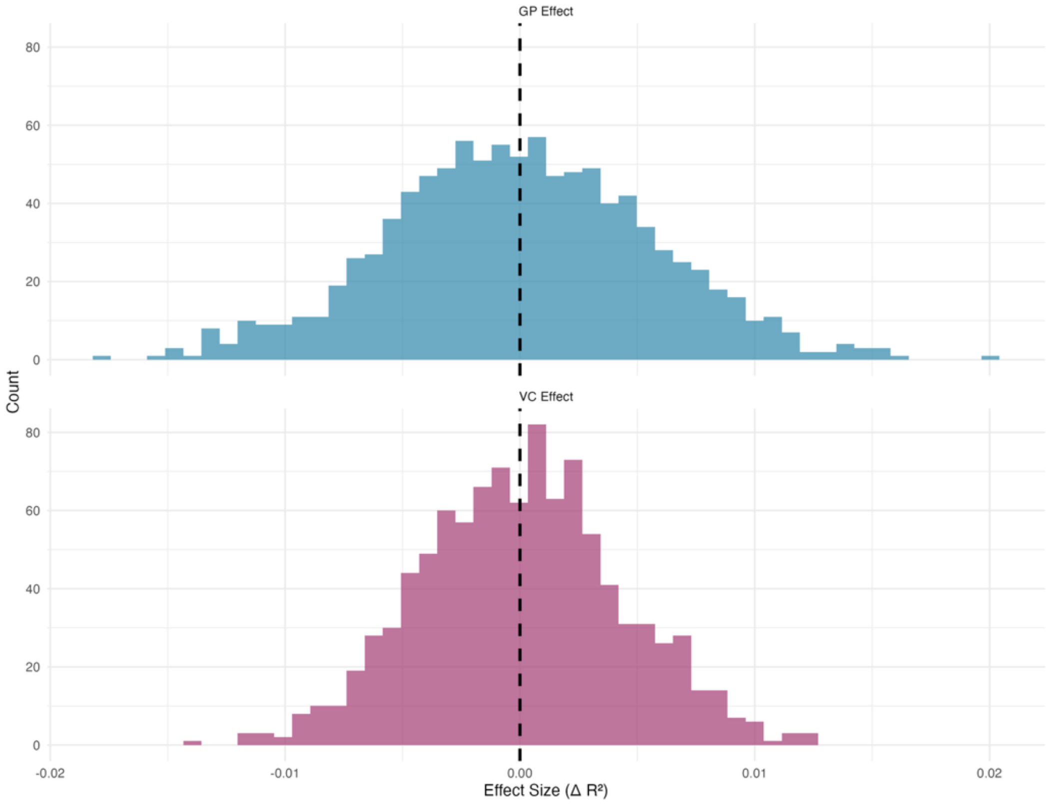
In this table, we present the results of the random forest algorithm concerning the out-of-sample R^2 prediction of our performance variable (i.e., IRR) for all three model specifications. Effect sizes represent absolute improvements in R^2. Confidence intervals and t-statistics from paired t-tests comparing model specifications. Significance: *** indicates $p < 0.001$.

Figure 1. Out-of-sample R^2 by Algorithm and Model Specification



This figure displays the out-of-sample R^2 values for all ML algorithms across three model specifications: baseline (i.e., fund characteristics and market conditions only), GP-enhanced (i.e., baseline with GP FE), and VC firm-enhanced (i.e., baseline with VC firm FE). Values are averaged across the 80 test periods of the time-series cross-validation procedure.

Figure 2. Wild Bootstrap Null Distributions of GP and VC Firm Effects



This figure shows the bootstrap distributions (with 1,000 resamples) generated under the null hypothesis of no effect. The dashed vertical line indicates zero. The observed GP effects and VC firm effects fall outside these null distributions, indicating statistical significance at conventional levels.

Appendix

A.1 Importance Analysis

To understand which features drive prediction when using the random forest algorithm, we examine feature importance measured by the percentage increase in mean squared error when each feature is randomly permuted. Table A1 presents the most predictive features, while Table A2 aggregates importance by feature category. *Mean Percentage Increase in MSE (Mean %IncMSE)* represents the average importance of features within each category, while *Total Relative Importance* sums the importance across all features in that category, computed as the number of features multiplied by their mean importance, which captures both the strength and breadth of each category's contribution.

Insert Tables A1 and A2 about here

The feature importance analysis reveals several interesting patterns. First, fund-level characteristics seem to be more important than GP (VC firm) FE (i.e., *Mean %IncMSE* = 16.7), with fund size and age as the two most predictive variables (i.e., *Mean %IncMSE* = 33.4 and *Mean %IncMSE* = 31.0 respectively). This aligns with economic intuition since larger, more mature funds face different deployment pressures and portfolio dynamics compared to smaller and younger funds. Second, temporal effects also show high importance, with vintage (i.e., *Mean %IncMSE* = 8.1, *Total Relative Importance* = 292.3) and year FE (i.e., *Mean %IncMSE* = 10.2, *Total Relative Importance* = 256.0) highlighting the role of market timing and fund cohort effects in determining returns. Unsurprisingly, the years around COVID-19 (i.e., 2021-2022) and the vintage years around the dot-com era (i.e., 1999-2001) display particularly high importance (i.e., *Mean %IncMSE* = 40.7 for funds with vintage year 2000 and *Mean %IncMSE* = 32.0 for funds with vintage year 2001), which reflect the impact of unpredictable events and market cycles on VC performance.

Individual GP FE reveal a pattern of distributed importance that provides support for our answer to the research question. While the most predictive individual GP shows a *%IncMSE* of 21.6, and the mean across all 811 GPs is 4.8, a deeper analysis of concentration reveals no outlier effect among GPs. When properly normalized, the top GP accounts for less than 1% (precisely, 0.7%) of total GP importance and the top 10% of GPs collectively account for only 23.7% of total GP importance. To achieve 50% of the cumulative GP effect would require considering 233 GPs (i.e., nearly 30% of all GPs), which demonstrates the breadth of the distribution of the GP effects. Despite this pattern, the total relative importance of GP FE (i.e., 3,909) exceeds any other feature category and represents approximately 84% of total feature importance across all categories. Based on the results presented in this section, GPs matter in aggregate through systematic contributions across many different managers, rather than through the performance of a few stars.

A.2 Robustness across ML Algorithms

To ensure that our findings are not algorithm-specific, we examine the GP and VC firm effects across all ML algorithms. Five out of seven algorithms exhibit a statistically significant GP effect using both our inference approaches. These algorithms are ridge regression, LASSO, elastic net, random forest and support vector machine. The other two algorithms, namely XGBoost and LightGBM, do not display significant results and have a poor baseline performance, which suggests that they are less suited for our prediction task. The pattern of results across algorithms strengthens confidence in our findings. All linear algorithms (i.e., ridge, LASSO and elastic net) show large percentage improvements in out-of-sample R^2 when GP FE are added, ranging from 66.3% to 76.9%. While these large improvements reflect the poor baseline performance of linear algorithms, they nonetheless indicate that GP FE provide substantial predictive value even when relationships are constrained to linearity. When looking at the non-linear algorithms the GP effect is weaker but statistically significant except for two cases (i.e., XGBoost and LightGBM). Similar results, albeit with smaller magnitudes, hold for our analyses on the VC firm effect.

Given that we test GP and VC firm effects across several different algorithms, we apply Benjamini-Hochberg false discovery rate corrections to control for multiple comparisons (Benjamini and Hochberg 1995). After false-discovery rate adjustment, the statistical significance of all the tests remains unchanged. In particular, GP effects remain significant after false-discovery rate correction for ridge, LASSO, elastic net, random forest and support vector machine using both wild bootstrap and t-test methods. VC firm effects similarly maintain significance across the same five algorithms when applying false-discovery rate correction. Only XGBoost and LightGBM fail to achieve significance both before and after the testing correction, consistent with their poor baseline performance in comparison with other non-linear ML algorithms. The robustness to testing corrections exhibited by our results provides additional confidence that the observed GP and VC firm effects represent real predictive improvements rather than false discoveries.

Appendix Figures and Tables

Table A1. Top 20 Most Important Features in Random Forest Algorithm

Variable	%IncMSE	Category
Vintage: 2000	40.75	Vintage Effect
Fund Size	33.42	Fund Characteristic
Year: 2021	32.68	Time Fixed Effect
Vintage: 2001	32.08	Vintage Effect
Fund Age	31.00	Fund Characteristic
Year: 2022	29.30	Time Fixed Effect
Number of LPs	25.70	Fund Characteristic
GP: 13464-37P	21.64	GP Identifier
Harvest Stage	21.68	Lifecycle Stage
GP: 13733-38P	21.53	GP Identifier
Year: 2000	19.99	Time Fixed Effect
GP: 13462-66P	19.19	GP Identifier
GP: 14037-22P	19.17	GP Identifier
GP: 12819-70P	18.62	GP Identifier
Vintage: 1999	18.57	Vintage Effect
GP: 12822-13P	18.46	GP Identifier
GP: 13595-77P	18.31	GP Identifier
GP: 13182-94P	17.61	GP Identifier
Vintage: 2006	17.42	Vintage Effect
Vintage: 1992	31.77	Vintage Effect

The table presents the top 20 features ranked by permutation importance (i.e., percentage increase in RMSE) for the random forest algorithm.

Table A2. Feature Importance by Category in Random Forest Algorithm

Variable	Nr of Features	Mean %IncMSE	Total Relative Importance
Fund Characteristics	7	16.75	117.30
Vintage Effects	36	8.12	292.30
Year Fixed Effects	25	10.24	256.00
Geographic Effects	14	5.10	71.40
Market Conditions	5	5.75	28.80
GP Identifiers	811	4.82	3909.00
Quarter Fixed Effects	4	-2.01	-8.00

The table shows feature importance by category for the random forest algorithm. Variables are grouped into categories; for each category we report the number of features, the mean permutation importance (i.e., %IncMSE) across features, and the total relative importance (i.e., mean times number of features).