Risk-Adjusted Returns of Private Equity Funds: A New Approach

ARTHUR KORTEWEG and STEFAN NAGEL*

ABSTRACT

This paper introduces a new metric, α, to benchmark the performance of individual private equity funds. Our metric is substantially less sensitive to noise in fund cash flows compared to the popular public market equivalent (PME) and its generalization (GPME), while having the same aggregate pricing implications as GPME. For a large data set of fund cash flows, α estimates have much lower standard deviation across funds than (G)PME. For buyout funds, PME and α are close, but deviate in certain subsamples. Using α increases power in regressions involving fund performance and improves performance predictability of future funds.

*Arthur Korteweg is from the University of Southern California Marshall School of Business, and Stefan Nagel is from the University of Chicago, NBER, and CEPR. We thank MSCI-Burgiss for providing the data, the Private Equity Research Consortium and the Institute for Private Capital for providing nonfinancial research support, and Wendy Hu for excellent research assistance. We thank Elise Gourier (discussant), Chris Schwarz (discussant), Juhani Linnainmaa, Luke Taylor and seminar participants at the 2016 Southern California Law and Finance Conference, the 2021 SFS Cavalcade, the 2024 Institute for Private Capital Spring Research Symposium at UNC, Boston College, Carnegie-Mellon, Dartmouth, UC San Diego, and the University of Melbourne for comments.

Allocations to private equity (PE) by U.S. pension funds and endowments have been rising in recent years.[1] For example, the 2020 NACUBO-TIAA Study of Endowments reports that university endowments invested 23% of total assets in venture capital and buyout funds, and another 6% in private equity real estate. The top 200 largest U.S. defined benefit public (corporate) pension funds in 2016 allocated an average of 9.0% (5.8%) of their assets to venture capital and leveraged buyout, and 8.3% (5%) to real estate.[2] The increased prominence of PE brings the problem of performance assessment to the forefront. Despite recent progress in assessing risk-adjusted returns at the asset class level, little is known about the properties of performance metrics when applied to individual PE funds. Assessing an individual fund's performance is, however, the typical problem in practice, as investors have to decide how to allocate money to managers. It is also an important issue for academic research on topics such as the degree of performance persistence and the relation between fund performance and characteristics of funds or managers.[3] This paper shows that the popular public market equivalent (PME) metric, and its generalized version, are very noisy measures of the risk-adjusted return to individual PE funds, and proposes a new measure that yields more accurate estimates.

Performance evaluation in PE is complicated because regularly observed returns are not available. A PE fund is organized as a limited partnership that formally lasts ten years. Lim-

[1]For the purposes of this paper, the term "private equity" encompasses all types of private equity, including but not limited to investments in leveraged buyout, venture capital, real estate, and natural resources.

[2]Numbers are from the Pensions and Investments annual survey of pension funds:
`http://www.pionline.com/article/20170206/PRINT/302069976/assets-of-top-funds-up-62-to-94-trillion`

[3]Papers that study performance persistence in private equity include Kaplan and Schoar (2005), Phalippou and Gottschalg (2009), Phalippou (2010), Chung (2012), Hochberg, Ljungqvist, and Vissing-Jorgensen (2014), Robinson and Sensoy (2016), Braun, Jenkinson, and Stoff (2017), Korteweg and Sorensen (2017), Nanda, Samila, and Sorenson (2020), and Harris, Jenkinson, Kaplan, and Stucke (2023). Examples of studies that relate PE fund performance to fund or manager characteristics include papers on returns to scale (e.g., Kaplan and Schoar, 2005, Metrick and Yasuda, 2010, Harris, Jenkinson, and Kaplan, 2014, Lopez-de-Silanes, Phalippou, and Gottschalg, 2015, Rossi, 2019), pay for performance (e.g., Chung, Sensoy, Stern, and Weisbach, 2012, Hüther, Robinson, Sievers, and Hartmann-Wendels, 2020), the relation between VC networks and performance (e.g., Hochberg, Ljungqvist, and Lu, 2007, Abell and Nisar, 2007, Liu and Chen, 2014), the relation between performance and LP identities (e.g., Lerner, Schoar, and Wongsunwai, 2007, Hochberg and Rauh, 2013, Sensoy, Wang, and Weisbach, 2014, Andonov, Hochberg, and Rauh, 2018, Cavagnaro, Sensoy, Wang, and Weisbach, 2019), the potential manipulation of interm performance during fundraising (e.g., Jenkinson, Sousa, and Stucke, 2013, Barber and Yasuda, 2017, Chakraborty and Ewens, 2018, Brown, Gredil, and Kaplan, 2019, Jenkinson, Landsman, Rountree, and Soonawalla, 2020, Hüther, 2023), and the relation between performance and idiosyncratic volatility (e.g., Ewens, Jones, and Rhodes-Kropf, 2013, Peters, 2018, Opp, 2019), amongst others.

ited partners (LPs) commit capital to the fund, which is called by the general partner (GP) when suitable investment opportunities are identified, and to pay management fees. When portfolio companies are sold, distributions are paid to the limited partners, after performance fees. The standard approach in the mutual and hedge fund industries that uses factor models to benchmark returns and estimate fund-level alphas, cannot be readily applied to a such a sequence of cash flows occurring at random times.

The generalized public market equivalent (GPME) approach of Korteweg and Nagel (2016) values the PE cash flow stream from an LP's perspective using a stochastic discount factor (SDF) that is calibrated to exactly price a set of benchmark assets (for example, public equities and risk-free bonds). If the PE cash flows can be replicated using a (possibly levered) investment in the benchmark assets, then the GPME is zero in expectation. This is a robust approach that accommodates the skewness and irregular horizon of PE cash flows, and it is well-suited to address the question of aggregate outperformance of the PE industry. As such, the measure works well in large data sets, taking averages over a large number of funds of varying vintages. However, for an *individual* fund, even if its expected GPME is zero, the realized value of the GPME may be far from zero. Typically, GPME realizations for individual funds are far too noisy to draw meaningful inferences about individual fund performance.

The root of the problem can be understood by analogy with factor model regressions for public market funds that have regularly observed realized fund returns. With observable returns, the typical exercise is to compare the realized fund return, R_t, to a benchmark portfolio return, R_t^b, which is a combination of factor portfolio returns scaled by their corresponding factor loadings. Subtracting R_t^b from R_t not only generates an excess return that is zero in expectation under the null of zero outperformance, but it also absorbs much of the period-by-period random variation in R_t that originates from unexpected shocks to the risk factors in R_t^b. As a consequence, $R_t - R_t^b$ delivers a relatively precise estimate of abnormal performance even for an individual fund. In contrast, discounting PE cash flows with the SDF in the GPME approach does not remove such common factor shocks to the same degree, which results in a much noisier estimate of abnormal performance.

In this paper we build on the GPME approach to develop a new benchmarking method that removes common factor shocks and is suitable for individual fund performance evaluation. To achieve this, the method requires two additional assumptions over and above the assumptions underlying the GPME method. First, we assume that fund cash flows and benchmark portfolio payoffs are jointly lognormal with constant covariance matrix and constant means. This is restrictive, but so are typical factor model regression approaches for public market funds which are essentially based on a mean-variance optimization framework. Second, we assume that betas (i.e., factor loadings) are the same across funds within an asset class (e.g., VC or buyout) and constant over time. This assumption is necessary to infer betas from data on aggregate performance. It could be relaxed by conditioning on fund characteristics and state variables, although this may not be feasible in practice given the small sizes of available data sets.

We construct a benchmark portfolio that matches the systematic risk of the PE cash flows. It takes continuously rebalanced levered positions in the SDF risk factors with weights given by factor betas. We show that if the benchmark portfolio uses the true betas as weights, then deflating PE cash flows with the return on this benchmark portfolio delivers exactly the same expected abnormal payoff as the GPME approach. In addition, deflating by the benchmark portfolio removes the common risk factor shocks from the PE cash flows, just as factor model regression in public market funds returns data do. As a consequence, we obtain a much less noisy measure of individual fund abnormal performance than the GPME. We use α as the label for the abnormal performance relative to the benchmark portfolio.

In practice we don't know the true betas. To estimate betas, we exploit the fact that deflating PE cash flows with the return on this benchmark portfolio must deliver the same expected abnormal payoff as the GPME approach. As a first step, we therefore estimate the GPME for a large group of funds and vintages within an asset class. We then apply the benchmark portfolio approach and search for the betas that deliver average abnormal payoffs that are exactly equal to the GPME and that minimize, at the same time, the cross-sectional variance of abnormal performance across funds. Finally, we evaluate individual fund performance with a benchmark portfolio that uses these estimated asset-class specific betas as factor weights.

Simulations confirm that α is overall a more accurate measure of fund-level abnormal PE return than GPME or the more restrictive public market equivalent (PME) of Kaplan and Schoar (2005) that deflates PE cash flows with the market return. There are only two cases where α is not strictly preferred. First, when beta is truly equal to one, then α and the abnormal performance measured by PME coincide, and the PME may perform slightly better in practice as it avoids the estimation of the parameters needed to compute the benchmark portfolio. Of course, in practice one would not know if beta is truly equal to one. Second, when betas are equal to the SDF loading on the corresponding factors (i.e., the factors' prices of risk), GPME and α coincide. We also show that these results are robust to reasonable violations of the main assumptions of constant betas and lognormality.

We use the MSCI-Burgiss Manager Universe, one of the largest samples of PE fund cash flow histories available, to take the benchmark portfolio approach to the data. Our sample includes a total of 1,630 VC funds and 1,073 leveraged buyout funds. The data are sourced exclusively from a large number of LPs, avoiding the natural biases introduced by sourcing data from GPs. The data are of very high quality because they are used for control purposes (audit and performance measurement) and cross-checked when multiple LPs invest in the same fund. In our baseline analysis, we use a CAPM SDF and, correspondingly, a benchmark portfolio with the market portfolio return as the only factor.

For venture capital funds, the differences between α and GPME at the individual fund level turn out to be modest. The across-fund standard deviation of α is 30% lower compared with the GPME. The differences are not big because the estimated VC beta of 2.37 is quite close to the estimate of the price of risk of the market factor in the SDF (3.61). As a consequence, discounting with the SDF or deflating with the benchmark portfolio produces relatively similar results. Put differently, with a market price of risk for the market factor in the vicinity of beta, discounting by the SDF removes common factor shocks almost as well as the deflating with the benchmark portfolio does.

In contrast, for the sample of leveraged buyout funds, we find much bigger differences between α and GPME. Here, the across-fund standard deviation of α is about 65% lower compared with the GPME. The reason is that the difference between beta (0.91) and market factor price of risk in the SDF (3.78) is much larger. This means that discounting by the SDF yields realized pricing errors that have large common factor components greatly magnify the noise and cross-sectional dispersion in fund-level GPME realizations compared with α.

We find different estimates of betas across early and late time periods, across fund size splits of the data, and for subclasses of venture capital. Therefore, the degree to which PME and GPME deviate from each other, and from α, varies over time and across fund characteristics. Although this may raise some concern regarding violations of the assumption of constant betas in the benchmark method, the simulations show that for the range of betas estimated in the data, the alpha metric is robust and more accurate than both PME and GPME.

When we repeat the analysis using two-factor models where we augment the market factor with an additional factor from the portfolios that underlie the Fama-French size and value factors, we find results that are similar. We again find much bigger differences between α and GPME at the individual fund level for buyout funds than for VC funds. Interestingly, VC funds load most strongly on the big-growth portfolio rather than on the small-growth portfolio that is often thought to be a good approximation of firms in VC's portfolios. Buyout seems to be best characterized by a tilt away from small-growth stocks.

Finally, we demonstrate the usefulness of the α performance measure in two applications that study the properties of PE funds' abnormal performance. First, we examine the relation between fund performance and fund size. The additional noise in the PME and GPME performance metrics substantially reduces statistical power to reject the null hypotheses in regressions of fund performance on fund size. Coefficient standard errors are as much as one-third to two-thirds lower when using α instead of GPME as the dependent variable, depending on the regression specification and asset class. Second, we look at performance persistence. Looking across different types of funds and different metrics of future performance, α overall performs better than (G)PME in predicting future fund performance. Only for VC funds GPME performs slightly better than α, but GPME does much worse for buyout funds. This is consistent with less noise contamination in α compared with (G)PME, and hence smaller errors-in-variables bias. The performance persistence analysis can also be viewed as a specification test. If the log-normal model underlying the α measure were strongly misspecified, the measure would not be useful in predicting future fund performance.

Our work relates to several other recent papers that propose advances in PE performance evaluation methodology. Several papers develop alternative approaches for individual fund performance evaluation. Gupta and Van Nieuwerburgh (2021) pursue an approach that is closely related to ours. Based on a multi-factor SDF and VAR dynamics for state variables that include valuation ratios and dividends of equity factors, they construct model-implied prices and cash flows of bond and equity strips (i.e., claims to single future cash flows at various horizons). Linear combinations of these strips then serve as benchmark portfolio prices and payoffs. A main difference to our approach is complexity. Their approach requires estimation of hundreds of parameters (the VAR companion matrix alone has more than 300, given there are 18 state variables) which is only feasible with ad-hoc parameter restrictions. In our case, with the CAPM SDF, we have four parameters: 2 for the SDF and 2 benchmark portfolio parameters. Considering this vastly different methodological complexity, it is interesting that our alpha estimates are quite close to theirs in their two-factor case not only in terms of mean but also cross-sectional dispersion across funds.[4]

Brown, Ghysels, and Gredil (2023) develop a state-space model to filter unobserved fund NAVs and estimate fund-level alphas and betas. Like Gupta and Van Nieuwerburgh, this approach also involves estimating many parameters and latent states over time. They estimate an average buyout fund beta of 0.97 to 1.15, depending on the model specification, which is only slightly higher than our estimate of 0.91. Their average VC fund has an estimated market beta of 1.39 to 1.60, somewhat lower than the 2.37 beta that we estimate. Our α estimates are generally not directly comparable to their estimates, since they estimate market-model intercepts, not CAPM alphas. However, for buyout funds, which have betas very close to one, the difference between those is minor. Like us, they find positive abnormal returns for buyout funds, but with larger magnitude.

Driessen, Lin, and Phalippou (2012) propose a heuristic approach that uses, like the

[4] In their table II, they report mean alphas (s.d.) of 0.28 (0.53) and -0.15 (1.36) for buyout and VC funds, respectively, while our corresponding estimates are 0.19 (0.56) and -0.18 (0.97).

PME method, a realized benchmark return to discount PE cash flows, but in their case the benchmark return is a beta-adjusted market return and includes an abnormal return parameter. Robinson and Sensoy (2013) use this approach, too, but they exogenously choose a beta instead of estimating it. In contrast, our approach is rigorously justified based on asset-pricing theory and our statistical assumptions. That said, the beta of 2.7 for VC funds that Driessen et al. estimate is very close to our estimate of 2.4.

Other recent papers propose other innovations in performance evaluation methodology, but without tackling the problem of individual fund performance evaluation. Gredil, Sorensen, and Waller (2020) use SDFs of habit-formation and long-run risks consumption-based asset pricing models. Using these SDFs, they find that venture funds have better risk-adjusted performance than with a CAPM SDF. Like the GPME method, their approach uses realized cross-products of SDFs and cash flows (albeit with an adjustment aimed at improving finite-sample performance of the estimator) and is hence not well-suited for evaluating individual fund performance. Ang, Chen, Goetzmann and Phalippou (2018) use Bayesian methods to extract from cash flow data and public markets factor returns an implied series of returns for groups of PE funds. This method is not applicable to individual funds, however. Boyer, Nadauld, Vorkink, and Weisbach (2023) construct an aggregate PE index from secondary market transactions of PE fund shares. Stafford (2022) construct a public markets replicating portfolio for PE buyout funds in aggregate.

The paper is structured as follows. Section I introduces theoretical background and our proposed method for estimating individual fund alphas. Section II discusses estimation. Section III introduces the data set. Section IV shows simulation results and robustness. Section V reports performance metrics estimated on the PE fund data. Section VI shows the impact of alpha on performance inference, and Section VII concludes.

I. Risk-adjusting Returns of Private Equity Funds

Our model setup is as follows. We take the perspective of the limited partners in a private equity fund. Let $\{C\} = \{C_1, C_2, .., C_K\}$ be the stream of (nonnegative) capital calls by the fund, and $\{X\} = \{X_1, X_2, .., X_J\}$ the (nonnegative) distributions to the LPs. For example,

$C = \$1$ means the LP sends one dollar to the GP, and $X = 2$ means the GP distributes \$2 back to the LP. See Figure 1 for an example of a fund with $K = 13$ capital calls and $J = 13$ distributions. To keep notation simple, we suppress fund-identifier subscripts in this section. Throughout the paper we use uppercase R for arithmetic returns, and lowercase r for log returns, that is, $r \equiv \log R$.

A. Generalized Public Market Equivalent

Korteweg and Nagel (2016) (KN) introduce a stochastic discount factor (SDF) approach to evaluating private equity fund outperformance. If an SDF M perfectly prices fund payoffs, then

$$E \sum_{j=1}^{J} M_{h(j)} X_j = E \sum_{k=1}^{K} M_{h(k)} C_k, \tag{1}$$

where $h(k)$ is the time since fund inception until the date of capital call k, and $h(j)$ accordingly for distributions. The SDF M_h is the one that prices payoffs over the relevant time horizon, h. KN assume an exponentially-affine CAPM SDF,

$$M_h = \exp(\delta h - \gamma r_h^m), \tag{2}$$

where r_h^m is the log return on the stock market portfolio from fund inception to h periods later. In analogy to the CAPM in its linear factor model version, this SDF evaluates payoffs from the viewpoint of an investor who is invested in the stock market portfolio and the risk-free asset. If an investment opportunity offers abnormal payoffs based on this SDF, this means that this investor would benefit at the margin from adding this asset to the portfolio in addition to the risk-free asset and the stock market portfolio.[5] Unlike the linear factor model version of the CAPM, the SDF in (2) can be applied to multi-period valuation problems. The multi-period SDF for a payoff at horizon h is simply the single-period SDF compounded over that horizon. Additional risk factors, such as those in Fama and French (1993), can be added easily, but to keep notation simple we focus on the single-factor case in laying out our

[5]Using this SDF in performance evaluation therefore does not require the assumption that the CAPM correctly prices the cross-section of stock returns. The SDF emerges from the first-order condition of the investor's optimal portfolio problem, without assumptions about equilibrium pricing.

approach.

KN estimate the SDF parameters δ and γ by requiring the SDF to perfectly price pseudo-funds that invest in T-bills and the stock market index. Then, if the private equity cash flow stream can be replicated by a (possibly levered) portfolio of the market index and the risk-free asset (or, more generally, the portfolios included in M_h), then (1) must hold with the SDF (2). Based on this, KN define the generalized public market equivalent (GPME) for a given fund as

$$GPME = \sum_{j=1}^{J} M_{h(j)} X_j - \sum_{k=1}^{K} M_{h(k)} C_k, \tag{3}$$

where the X_j and C_k are distributions and capital calls scaled by the sum of capital calls discounted with the risk-free rate. These scaled distribution cash flows, $\{X\}$, can then be interpreted as the dividends from an investment of approximately \$1 in the fund, and the expected GPME can be interpreted as the NPV of that investment. If the capital call cash flows are deterministic, this interpretation becomes exact. If there are no abnormal returns to investing in PE, then the *expected* GPME equals zero.

The GPME generalizes the well-known public market equivalent (PME) metric of Kaplan and Schoar (2005), which is a special case of equation (2) with $\delta = 0$ and $\gamma = 1$ (see Sorensen and Jagannathan, 2015, and Korteweg and Nagel, 2016). We use a slightly different definition of PME than the original Kaplan and Schoar paper, which uses the ratio of discounted distributions and capital calls. Our definition uses the difference, which avoids some Jensen inequality issues with the ratio, as explained in Korteweg and Nagel (2016). In the Kaplan-Schoar paper, a fund that can be replicated by the market has a PME of one, while according to our definition that fund's PME is zero.

B. *Benchmark portfolio for individual fund returns*

GPME works well in expectation, that is, when estimated across a large number of funds that cover different time periods. However, a typical problem both in academia and in practice is to assess the performance of an individual fund. In simulations below we show that the GPME, though statistically consistent, can be a very noisy measure of a single fund's outperformance. In this section we develop a benchmark portfolio comparison method that works equally well in expectation, but yields more accurate estimates of fund-level abnormal returns.

For this purpose, we need to impose a little more structure on the problem than the GPME framework does. This gives up some of the generality of the GPME approach, but it enhances statistical power. Specifically, we add the assumption that market factor returns and fund payoffs are IID over time and distributed jointly lognormal. This means that the joint distribution that the log market return $r^m_{h(j)}$ and log payoff x_j realized between time t and $t+h(j)$ are drawn from is the same bivariate normal distribution for all t. Henceforth we drop the j index in $h(j)$ to reduce notational clutter. We further assume a constant risk-free rate r_f. Finally, we assume that the market beta of log payoffs, $\beta = cov(x_j, r^m_h)/var(r^m_h)$, is horizon-independent and does not vary across funds within an asset class. Defining $\mu = \log E[R^m_1] - r_f$ and $\sigma^2 = var(r^m_1)$, these assumptions imply

$$\begin{pmatrix} r^m_h \\ x_j \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu h + r_f h - \frac{h}{2}\sigma^2 \\ E[x_j] \end{pmatrix}, \begin{pmatrix} \sigma^2 h & \beta\sigma^2 h \\ \beta\sigma^2 h & var(x_j) \end{pmatrix}\right]. \tag{4}$$

Since h-period log market returns are the sum of single-period market returns, expected return and variance scale linearly with h. For generality of the approach, we do not restrict the x_j payoff to have a representation as a sum of $h(j)$ single-period IID random variable realizations. Therefore, $E[x_j]$ and $var(x_j)$ do not necessarily scale with h (but the component of x_j spanned by r^m_h scales with h and the definition of β implies that $var(x_j) \geq \beta^2\sigma^2 h$ holds).[6]

Under these assumptions, the GPME method with SDF (2) corresponds to benchmarking PE returns with a log-linear version of the CAPM,

$$\log E[R_h] - r^f_h = \beta\left(\log E\left[R^m_h\right] - r^f_h\right), \tag{5}$$

[6]Consider the following extreme and unrealistic example to illustrate the generality of the assumption regarding x_j: Suppose that shortly after a fund is set up, an idiosyncratic normal shock u is drawn, and then all fund distributions are equal to u, i.e., $x_j = u$ for any j. This means that uncertainty about the fund cash flows is resolved once the single shock u is realized, and the fund beta is zero. In this case, the variance $var(x_j)$ is equal for all j. In contrast, if x_j at horizon $h(j)$ was generated as a sum of $h(j)$ normal random shocks with variance σ^2_x, then $var(x_j) = \sigma^2 h$, i.e. variance would grow with horizon $h(j)$. Our method accommodates both data-generating processes (and others, as long as x_j is jointly normal with log market returns).

where $r_h^f = r_f h$. To see this, note that the SDF parameters are pinned down by the requirement that the SDF correctly prices the risk-free asset and the market portfolio of public equities. The pricing equations for these two assets can be solved for

$$\begin{aligned} \delta &= -r_f - \frac{1}{2}\gamma^2\sigma^2 + \gamma(r_f + \mu - \frac{1}{2}\sigma^2), \\ \gamma &= \frac{\mu}{\sigma^2}. \end{aligned} \tag{6}$$

Evaluating the log of the pricing equation $E[M_h R_h] = 1$ with these parameter values then yields (5).

Even though PE fund returns are not observable, we *can* benchmark fund cash flows under the statistical assumptions we made above. Consider the valuation of a single cash flow, X_j, in (3). Using the SDF (2) with the parameter values in (6), and taking expectations, we obtain

$$E[M_h X_j] = E\left[\exp\left\{-r_h^f - \beta(r_h^m - r_h^f) + \frac{h}{2}\beta(\beta-1)\sigma^2\right\} X_j\right]. \tag{7}$$

(PROOF: See appendix A.)

Since β is horizon-independent, equation (7) applies to the valuation of every other cash flow in the summations in (3), using the same β. Therefore, we get

$$E[GPME] = \left(E\sum_{j=1}^{J} \frac{X_j}{R_{h(j)}^b}\right) - \left(E\sum_{k=1}^{K} \frac{C_k}{R_{h(k)}^b}\right), \tag{8}$$

with the benchmark portfolio return

$$R_h^b = \exp\left\{r_h^f + \beta(r_h^m - r_h^f) - \frac{h}{2}\beta(\beta-1)\sigma^2\right\}. \tag{9}$$

Thus, if we deflate each cash flow X_j by the realized gross return that the benchmark portfolio would have earned over the horizon $h(j)$ from initial investment into the fund until realization of X_j, we get the same abnormal payoff in expectation, $E[GPME]$, as by discounting with the SDF in (3).

The return R_h^b can be interpreted as the return on a levered portfolio that matches the systematic risk of the fund cash flows. More precisely, if time is continuous, R_h^b is the h-period return on a continuously rebalanced portfolio that, at every instant, invests a share β in the market portfolio and $1-\beta$ in the risk-free asset (alternatively, one can view R_h^b as the exponential of a second-order Taylor approximation of the log return of a portfolio that invests a share β in the market portfolio and $1-\beta$ in the risk-free asset at the beginning of the return measurement interval; see Campbell and Viceira (2002), Chapter 2).

While deflating cash flows with R_h^b delivers the same expected abnormal payoff as discounting with the SDF does, it is important to note that $1/R_h^b$ is not an SDF, as it depends on an asset-specific parameter (β).[7] Moreover, while expected values are the same, discounting with the SDF and deflating with the benchmark portfolio produces a different realized abnormal payoffs for an individual fund. Define the fund-level abnormal performance metric

$$\alpha = \sum_{j=1}^{J} \frac{X_j}{R_{h(j)}^b} - \sum_{k=1}^{K} \frac{C_k}{R_{h(k)}^b}. \tag{10}$$

Compared to (3), α is a more accurate measure of abnormal fund-level performance, because the benchmark portfolio eliminates variation in payoffs due to the factors in the benchmark portfolio. To see this, consider again a single cash flow X_j. In logs, the cash flow deflated by the benchmark portfolio return is $x_j - r_h^b$ while the realized product of the cash flow with the SDF is $x_j + m_h$. Now note that

$$x_j - E[x_j] = \beta(r_h^m - E[r_h^m]) + \epsilon_j, \tag{11}$$

$$r_h^b - E[r_h^b] = \beta(r_h^m - E[r_h^m]), \tag{12}$$

$$m_h - E[m_h] = -\gamma(r_h^m - E[r_h^m]), \tag{13}$$

where ϵ_j is an idiosyncratic shock. Therefore, after deflating by R_h^b, the unexpected log payoff

[7] In contrast, the reciprocal of the SDF *is* a proper portfolio return: the return on the growth-optimal portfolio, as shown in Long (1990). Under the assumption that the SDF is (2), the growth-optimal portfolio is the return on a continuously rebalanced portfolio that, at every instant, invests a share γ in the market portfolio and $1-\gamma$ in the risk-free asset. In the special case of a log-utility SDF, $\gamma = 1$, the growth-optimal portfolio is the market portfolio, and we obtain the PME metric.

is

$$x_j - E[x_j] - (r_h^b - E[r_h^b]) = \epsilon_j. \tag{14}$$

By virtue of β being a regression coefficient in a regression of x_j on r_h^m, deflation with R^b completely removes the component of x_j related to r_h^m and only the idiosyncratic part ϵ_j remains. In contrast, the unexpected log product with the SDF is

$$x_j - E[x_j] + (m_h - E[m_h]) = (\beta - \gamma)(r_h^m - E[r_h^m]) + \epsilon_j. \tag{15}$$

Unless it happens to be the case that $\beta = \gamma$, the SDF based measure in (3) is therefore subject to an additional source of noise coming from the $r_h^m - E[r_h^m]$ component. Especially when $|\beta - \gamma|$ is big or when the realized average market portfolio return during a fund's lifetime was far from its expected value, the SDF-based GPME can deviate far from α.

Thus, while the SDF approach provides a valid measure of abnormal performance in the sense that it is correct on average in a large-T sample, the benchmark portfolio approach, while also correct on average, should minimize the noise in performance assessment and therefore provide a more accurate measure of abnormal performance suitable for benchmarking individual funds.

C. Specification choices

To apply this method, an investor evaluating fund performance needs to choose risk factors to include in the SDF, and hence in the benchmark portfolio, and the set of funds to be used for the estimation of factor betas. As noted in the description of the GPME method above, the basic principle is that the SDF should reflect the investor's existing portfolio. For example, an investor currently invested in U.S. public equity and U.S. Treasury bills should use a U.S. public equity index returns as risk factor. The abnormal performance metric then tells the investor whether payoffs of the fund under consideration can be replicated by a U.S. public equity index investment or whether investing in the fund would improve the risk-return tradeoff faced by the investor (see also Korteweg, Panageas, and Systla, 2024, for more details and an SDF specification that uses the return on an investor's own wealth portfolio).

Adding additional factors can further be useful to find public markets replication strategies and to improve statistical power. For example, a U.S. investor currently invested in U.S. public equity and evaluating a European PE fund, might also consider a potential investment in European public equity as an alternative to the European PE investment. In this case, the investor can add a European public equity index to the SDF and the benchmark portfolio, alongside the U.S. public equity index. The analysis will then reveal whether an investment in the European PE fund can be replicated by taking on exposure to European public equity. Since adding the European public equity index as risk factor will absorb some of the common factor components of the European PE fund payoffs, the resulting estimate of abnormal performance should also be more precise than in the single-factor case.

Finally, our method requires an assumption that the factor betas of the fund being evaluated is the same as the factor beta of the funds that are used for estimation of beta. Therefore, as far as possible while maintaining a sufficiently large sample size, and to the extent that there is material heterogeneity in betas, it may be beneficial to choose a set of funds that is similar in terms of the type of firms the funds invest in, such as their life-cycle stage, industry, or geographical location.

II. Estimation

The previous section ignores estimation of β and the SDF parameters, δ and γ. Estimation is not a trivial task because PE fund returns are unobserved in practice, which materially complicates the estimation of β compared to, say, mutual funds applications. To make estimation feasible, we impose the assumption that β is identical within a group of PE funds. As we explain below, this allows us to use information from the cross-section of fund pricing errors to pin down β. The performance of the GPME and alpha measures, both in absolute and relative terms, will depend on the degree of estimation error in these parameters. The performance of alpha will also depend on how reasonable the additional assumptions of homogeneous betas and lognormality are in the data, and the robustness of the approach to violations of these assumptions. We analyze these issues in the simulation section below.

As a first step, we estimate the parameters δ and γ in the SDF (2) as discussed in Korteweg and Nagel (2016), by fitting the SDF to artificial PE funds that invest only in Treasury Bills and the stock market portfolio. We then apply this SDF to calculate the expected GPME (a single number for the whole sample of funds), again exactly as in KN. We label this asset-class level abnormal performance as $\bar{\alpha}$.

The next step is to estimate β. To incorporate a degree of robustness to the lognormality assumption, we use σ_h^2, the empirical variance of factor returns over horizon h, instead of $h\sigma^2$ in the Jensen inequality adjustment.

$$R_h^b = \exp\left\{r_h^f + \beta(r_h^m - r_h^f) - \frac{1}{2}\beta(\beta-1)\sigma_h^2\right\}. \tag{16}$$

We then estimate β by exploiting the fact that deflating with R_h^b should produce the same pricing error $\bar{\alpha}$ that we get from discounting with the SDF. One might therefore think that to estimate β, one could remove $\bar{\alpha}$ from the right and left-hand side of (10), averaged over all a funds within an asset class, and then simply solve this equation for β. Intuitively, this approach looks for the degree of leverage that the benchmark portfolio needs to have to produce a pricing error equal to $E[GPME]$. However, this equation has two solutions. Note that $1/R_h^b$ and M_h both have the functional form $\exp(ah - br_h^m)$ with $b = \beta$ in the case of R_h^b and $b = \gamma$ for the SDF. Since both deflating with R_h^b and discounting with M_h produces the same pricing error $\bar{\alpha}$, both yield an expected value of (10) of zero after $\bar{\alpha}$ has been removed from both sides of this equation. In other words, one of the solutions just reproduces the SDF that was used to obtain $\bar{\alpha}$.

To find the solution with $b = \beta$ rather than $b = \gamma$, we exploit the fact that the solution with $b = \beta$ should produce smaller cross-sectional dispersion in abnormal performance across funds than the solution with $b = \gamma$. The reason is that abnormal performance based on the SDF method with $b = \gamma$ includes the additional component $(\beta - \gamma)(r_h^m - E[r_h^m])$ in (15). Funds that do not have completely identical lifetimes and cash-flow timing will have different realizations for this component in addition to the differences that arise from the idiosyncratic component ϵ_j. In contrast, the benchmark portfolio method abnormal performance measure in (14) has only the ϵ_j term, and hence smaller cross-sectional dispersion.

Using a data set of $i = 1 \dots N$ funds, we therefore estimate

$$\widehat{\beta} = \arg\min_{\beta} Q(\beta) \text{ with } Q(\beta) \equiv \frac{1}{N}\sum_{i=1}^{N}(\alpha_i - \bar{\alpha})^2$$
$$\text{subject to } \frac{1}{N}\sum_{i=1}^{N}(\alpha_i - \bar{\alpha}) = 0 \tag{17}$$

where α_i is a function of β and $\bar{\alpha}$ is given from the SDF-based $E[GPME]$ estimation. Thus, among the two solutions that satisfy the pricing constraint, we look for the one that minimizes the cross-sectional dispersion in abnormal performance.

A. Standard error estimation

The methodology outlined above allows researchers and practitioners to construct point estimates of fund-level abnormal performance. Estimation of standard errors for fund-level abnormal performance measures is far more difficult. The alpha estimation error has two sources: Estimation error in the factor loadings (betas) and estimation error in the factor-adjusted payoff (given known factor loadings).

Estimating a standard error for factor loadings is feasible. GMM estimation of the GPME provides a standard error for the expected GPME on the left-hand side of equation (8). Applying the delta method, one can then numerically solve for the standard error of the factor loadings implied by the standard error of the expected GPME.

However, the far bigger contribution to the standard error of fund-level alphas likely comes from the second component. Estimating this contribution is more difficult because, effectively, we only have one observation of the factor-adjusted payoff for each fund. For individual public markets funds, one can infer the standard error of abnormal performance from the time-variation in observed periodic returns during the life-time of the fund. In contrast, in private markets, with cash flow data, we only have a single observation of abnormal payoff per fund. This is analogous, in the public markets case, to only having a single long-term return observation over each fund's lifetime, but not the higher-frequency periodic return observations. We are not aware of any existing methods in the literature that provide estimates of standard errors of fund-level abnormal performance measures that are based on cash flow data, including, for example, fund-level estimates of PME standard errors.

We briefly sketch a simulation approach here that could be used to estimate the standard error. In Appendix C, we run Monte Carlo simulations with simulated cash flows under the null hypothesis of no abnormal performance. The parameters of the data-generating process are set such that the cash flows have broadly realistic empirical properties. In these simulations, since the null hypothesis is imposed, all deviations of estimated fund-level alphas from the true alpha of zero are due to estimation error. Thus, the standard error of an individual fund's alpha can be inferred as the square root of the sum of the cross-sectional variance of estimated alphas and the squared mean of estimated alphas. For example, in the case of true $\beta = 2$ in the small-sample simulations in Table C.II, this calculation yields a large standard error of $[0.523^2 + (-0.042)^2]^{\frac{1}{2}} \approx 0.525$. We also find that beta estimation error contributes very little to this standard error because the root mean squared deviation between the estimated alphas (based on estimated betas) and the true realized alpha (based on the true beta), is only 0.126.

It is plausible that the standard errors are so large. For example, for a public markets fund with concentrated exposure to only a few stocks and a resulting high factor-adjusted volatility of 20% per year, the 10-year abnormal return of this individual fund would have a standard error of roughly $0.20 \times \sqrt{10} \approx 0.63$, i.e., roughly the same magnitude as what we calculated for the simulated funds.

While a simulation approach along these lines seems useful for tackling the standard error question, a more in-depth investigation of this problem is needed, and makes for an interesting area for future research.

III. Data

We use private equity fund cash flow data provided by MSCI-Burgiss, a global provider of investment decision support tools for investors in private capital markets. The data are sourced exclusively from hundreds of LPs, including public and private pension funds, endowments and foundations. We have the complete cash flow (net of management fees and carried interest (profit shares) paid to the GPs) and valuation history of all funds that the participating LPs invested in. The data are accurate and up-to-date, because they are used for audit, performance measurement and reporting purposes, and they are cross-checked when different LPs invest in the same fund. The market return and risk-free rate data are from Ken French's data library.[8]

We separate our sample of 2,703 funds into 1,630 VC funds and 1,073 buyout funds, representing all U.S. funds raised before 2017 that have at least $5 million in committed capital in 1990 dollars (following Kaplan and Schoar, 2005). The cash flow data run through the first quarter of 2022, so we have at least five years of cash flow data for all funds in our sample.

Table I Panel A reports that the average (median) VC fund has $241 million ($150 million) in committed capital, and involves 38.8 (35.0) net cash flows spanning 13.6 years (14.2 years). There are 608 unique VC firms in the sample, with an average (median) of 2.7 (2.0) funds per GP.

The average (median) buyout fund has committed capital of $1,116 million ($441 million), runs for 12.7 (13.0) years, comprising 55.8 (50.0) cash flows. The sample represents 436 unique GPs, with 2.5 (2.0) funds per GP.

As of the first quarter of 2022, 45% of VC funds and 40% of buyout funds were fully liquidated. These are funds that are at least 10 years old and have zero remaining net asset value (NAV). Some funds run longer than 10 years but have small net asset balances, so we also consider a less strict definition that considers funds to be liquidated if they are over 10 years old and their most recent reported NAV is less than 5% of committed capital. By this "95% liquidated" measure, 53% of VC funds and 51% of buyout funds are liquidated.

Performance of private equity funds is typically measured by their internal rate of return (IRR), and their total value to paid-in capital (TVPI) multiple. The TVPI is defined as the total distributions plus the NAV of unrealized investments, divided by total capital contributions to the fund by LPs. Both measures are net of fees and carried interest paid to the GP. The IRR of the average (median) VC fund is 17.2% (11.3%) per year, compared to 15.9%

[8]We thank Ken French for providing the data on his website.

(15.1%) for buyout. Weighted by fund size, the average IRRs are 15.7% and 15.6% for VC and buyout, respectively. The average (median) TVPI is 2.6 (1.7) for VC funds, and 1.9 (1.8) for buyout funds. The size-weighted TVPIs are 2.5 and 1.8, respectively.

Panel B of Table I reports the distribution of the number of funds and their performance metrics by vintage year (defined as the year of the fund's first capital call). IRRs and TVPIs appear cyclical for both venture capital and buyout. Venture capital experienced high returns in the early to mid-1990s vintages, followed by a decade of poor returns. Performance has recovered since the end of the global financial crisis. Buyout returns have been relatively more stable, but with some slumps in the late 1990s and around 2005. Note that funds raised after 2012 have not yet reached their ten-year lifetimes by the first quarter of 2022, which is the end of our sample period, and performance statistics depend increasingly on unrealized NAV toward the end of the sample. IRR and TVPI for fully liquidated funds tend to be higher, because many cash distributions occur towards the end of a fund's life, and in many cases GPs are conservative in updating NAVs (Jenkinson, Sousa, and Stucke (2013) and Brown, Gredil, and Kaplan (2019)).[9]

Harris, Jenkinson, and Kaplan (2014) perform a careful comparison of MSCI-Burgiss with commonly used private equity data sets (including Preqin, Venture Economics, and Cambridge Associates). They find that compared to Preqin, MSCI-Burgiss contains considerably more funds with cash flow histories, which are necessary to compute the performance metrics in this study. In their sample, MSCI-Burgiss' coverage of VC funds is less extensive than Venture Economics and Cambridge Associates in the early years, but increases significantly over the sample period. For buyout, they find that MSCI-Burgiss offers excellent coverage since 2000, but contains relatively few funds before 1993.

In terms of performance, Harris et al. report that MSCI-Burgiss yields qualitatively similar results to the Cambridge Associates and Preqin data, whereas Venture Economics data are downward biased (see also Stucke, 2011). Crucially, performance does not appear to be biased by the voluntary approval of LPs to allow their data to be included by MSCI-

[9]Since 2008, PE firms (both venture capital and buyout) are required by the Financial Accounting Standards Board under ASC Topic 820 (formerly known as FAS 157) to value their assets at fair value each quarter. While valuations have become more accurate since (Jenkinson et al., 2020, Easton et al., 2021), they remain conservative (Brown et al., 2019).

Burgiss.

Compared to Harris et al.'s sample, our MSCI-Burgiss data includes eight more years of data (their sample ends with the 2008 vintage year funds). More importantly, many more LPs have allowed their data to be included since the conclusion of their study, resulting in additional fund data going back all the way to the earliest vintages. Consequently, our sample of VC and buyout funds is about twice as large as the MSCI-Burgiss sample in Harris et al. (2014).

IV. Bootstrap Simulations

Our method rests on the key assumptions that cash flows are generated from IID log-normal distributions and that betas are homogeneous within the group of funds that we apply the method to. These assumptions are unlikely to hold exactly in the data. Therefore, before applying the method, we want to check whether violations of these assumptions could be severe enough in a realistic setting to substantially bias the estimates of α and β. To assess these statistical properties, we simulate a data set of PE funds of the same sample size as the fund data introduced in the previous section, and with cash flows that approximate the timing and distribution of these cash flows, but with known α and β, and without a log-normality assumption imposed on cash flows.

In the mutual funds literature, a common approach to simulate fund returns, following Fama and French (2010), is to remove estimated alphas from fund returns and then draw bootstrap samples from the resulting adjusted returns. This approach relies on the presumption that the estimated alphas are unbiased, which a reasonable assumption in a linear factor model regression setting. In contrast, in our setting here, we do not want to assume that alpha estimates obtained with our method are unbiased. Instead, we would like the simulations to reveal whether misspecification of the distributional assumptions in our method could lead to biases.

For this reason, we use a different approach. First, we divide each sample fund's cash flows by the size of the fund, such that commitment sizes are comparable across funds. Next, we construct a set of artificial funds by taking the sample funds' capital calls and investing them in the levered CRSP value-weighted stock market index, financed at the T-bill yield. We choose a β to control the degree of leverage and rebalance monthly to ensure that leverage remains constant. Distributions occur on the same dates as the corresponding PE fund, and are a fraction of the remaining capital in the fund, computed following the algorithm in KN Section I.A. The result is a data set of cash flows for 2,703 funds with known betas, and zero risk-adjusted outperformance relative to the (levered) public market equity index. Appendix B shows that the cash flow distribution of these bootstrapped funds is very similar to the observed funds. As final step, we generate a realistic cross-sectional distribution of true α by adding an additional cash flow to the first capital call of each fund, drawn randomly (with replacement) from the empirical distribution of the sample funds' PMEs (where PME is defined as the sum of fund cash flows discounted by the market return, as in Section I.A above). These additional cash flows are the true α of the bootstrapped funds. We can then judge an estimator of α by the distance between the true α and the estimated α.

Table II summarizes estimated alphas and (G)PMEs for the bootstrapped funds when all funds share the same true beta. We show results for various levels of the common β, ranging from 0.5 to 3, in increments of 0.5. The block of columns on the left shows the means and standard deviations of of the raw estimates of α, GPME, and PME. In the case of $\beta = 1$, the PME is the correct measure of abnormal performance and the mean (standard deviation) of PME of 0.242 (1.187) represents the distribution of the true realized alphas that we generated in our bootstrap. In this $\beta = 1$ case, the α estimated with our method recovers exactly the same mean and standard deviation.

The block of columns on the right shows means and standard deviations of α, GPME, and PME in excess of the true realized α. The mean excess alpha shows that, on average, as long as β is not too high, the estimated α does not have much bias since the mean excess α is very close to zero. Only for $\beta > 2$ do we get a substantial negative bias in the estimated α. As the table shows, the reason is that for β above this threshold, we substantially underestimate β. Importantly, however, the standard deviations of the excess alphas are much smaller than the standard deviations of GPME, even when true β is as high as 2.5. As a consequence, the RMSE shown in Panel B is much smaller for the α metric than for the GPME. This means that even though the α estimates are biased when $\beta = 2.5$, they are still much closer to the true alphas than the GPME (or the PME).[10]

The RMSE of GPME in Panel B decreases as beta rises. This is expected. The γ parameter in the SDF is estimated to be 3.674. In the special case $\beta = \gamma$, $1/R_h^b$ and M_h are identical and fund-level GPME and α estimates coincide. Thus, as β approaches the estimated γ, the RMSE of the GPME gets closer to the RMSE of the α estimates. In fact, the RMSE of the GPME already catches up with the RMSE of α estimates at β somewhat lower than 3.674 because the underestimation of β in our method introduces additional error.

Overall, the simulations demonstrate that α is a substantially more accurate measure of fund abnormal performance than (G)PME, with the exception of very high levels of true beta, where alpha and (G)PME are equally accurate.

Next, we relax the assumption that all funds have the same true beta. Table III shows the results of repeating our bootstrap exercise when we draw each fund's beta from a normal distribution centered around a true mean (ranging from 0.5 to 3) with a standard deviation of 0.25. For example, for a true mean beta of 1, this implies that 95% of funds have a beta between 0.5 and 1.5. The alpha estimates suffer somewhat relative to the constant beta case, especially for lower betas, but remain quite robust overall. For example, the RMSE for alpha slips from 0.094 for a constant beta of 2 to 0.111 in the case of heterogeneous betas centered around the same mean. The RMSE remains far lower than for GPME and PME. At all levels of true mean beta, alpha remains the dominant measure of fund-level outperformance, compared to (G)PME.

V. Fund-level Abnormal Returns

Table IV reports key statistics of the risk-adjusted fund performance metrics (α, GPME, and PME) as well as the parameter estimates of the SDF and the common fund beta, for ven-

[10]The underestimation of beta could be a small-sample bias or a consequence of violations of the IID log-normality assumption that our method relies on. We are not able to completely resolve this question, but simulations we report in Appendix C are suggestive that the distributional assumptions could be the reason. In these alternative simulations, we generate cash-flow data from an IID log-normal distribution and we find that beta estimates are very close to true betas, even at high levels of true beta, and in small and large samples. So it does not seem to be a small-sample problem.

ture capital and buyout separately. The distribution of fund performance differs substantially across the three metrics.

Focusing first on venture capital, the mean GPME (and hence, mean α) is negative but insignificant at -0.184.[11] While similarly statistically insignificant, this result is lower than the estimate reported by KN, who estimate a mean GPME of -0.103 in a shorter sample of Preqin VC funds that includes vintages up to 2008. In contrast, the mean PME in Table IV is 0.258, which is positive and significant (KN estimate a lower PME of 0.048). PME overstates the risk-adjusted return because the beta of VC is estimated at 2.4, far above the value of one implicitly assumed in the PME calculation, resulting in a PME-implied expected return that is too low.[12] The literature reports VC fund beta estimates between 1.0 and 2.8 (see Korteweg, 2019, 2023, for reviews of the literature). The wide range of estimated betas is due to differences in data sources, sample periods, and methodologies, but it's reassuring that our estimate falls within that range.

Despite the difference in central location of the GPME and PME distributions, the level of dispersion around the mean is very similar. Both measures have standard deviations around 1.4. In contrast, the standard deviation of fund-level α's is about 30% lower, at around 1.0. That the standard deviation of α is lower is consistent with α being a more precise measure of abnormal performance, in line with our theory and simulations. Specifically, our simulations in Table II suggest that α should produce lower RMSE than GPME. Since GPME and α have, by construction, the same mean, lower RMSE should manifest in terms of lower standard deviation. The differences in standard deviations in Table IV can therefore be viewed as a specification test. That the prediction of lower standard deviation is borne out in the data suggests that the assumptions underlying our theory and simulations are reasonably well specified.

The estimated beta around 2.4 also helps understand how the differences in dispersion of performance measures arise. Since beta is neither equal to the value of $\gamma = 1$ assumed in

[11] Since mean alpha and mean GPME coincide by construction, the p-value for the test of the null hypothesis that mean alpha is zero is identical to that of the GPME, so Table IV does not separately report it

[12] A different but equivalent perspective of the difference between PME and GPME is that PME assumes that the SDF is the reciprocal of the return on wealth, such that the SDF parameters δ and γ are zero and one, respectively. However, as Table IV shows, the estimate of γ in the VC sample is far from one.

the PME nor the estimate of $\gamma = 3.6$ used in the GPME calculation, PME and GPME are contaminated with additional noise at the individual fund level that is eliminated in the α calculation.

To illustrate the comparison between PME, GPME and α more clearly, Figure 2 plots the histograms of these three performance metrics. There are far fewer funds with extreme outcomes than the (G)PME metrics suggest. Many funds have GPMEs below -2 or above 2, whereas almost all fund α's are confined with this range. The PME distributions are visually closer to α, but are also more dispersed, as is most clearly seen from the lower peaks of the distributions at the mode and the higher incidence of outliers. Figure 3 show the time-series of the cross-sectional mean and median of the performance metrics. Compared to alpha, PME shows more time-series volatility around the dot-com boom and bust, whereas GPME had a stronger dip during the global financial crisis. Below we will analyze the implications of these differences for the evidence on performance persistence, and the relation between performance and fund characteristics.

Turning to the buyout results, we find that the mean GPME of 0.190 is above the mean PME of 0.161. The means of the two metrics are considerably closer than in VC, because the estimated buyout beta is 0.9, not so far from the PME assumption that beta equals one. This estimate is at the lower end of the 0.7 to 2.7 range of buyout fund betas reported in the literature, based on a variety of methodologies, data sources, and sample periods (see Korteweg, 2019, 2023). Since the beta estimate is below one, the mean GPME is above the mean PME, unlike in venture capital where a beta above one resulted in a lower mean GPME.

It might appear puzzling that despite being higher, the mean GPME is not statistically different from zero whereas the mean PME is highly significant. The reason is that the GPME requires estimation of the SDF parameters, whereas PME assumes they are fixed and given. Relaxing the PME restrictions raises the standard error of the estimated GPME.

Comparing α and PME, the difference between the two distributions is considerably smaller in buyout than in VC. Though α has a higher mean and median, the standard deviation, skewness, and kurtosis of the two distributions are close. The reason is again that the estimated beta is not far from $\beta = 1$, as assumed in the PME calculation. In the full sample of buyout funds considered here, PME and α thus look similar. However, the simulations suggest that α yields more accurate fund-level results when beta is not exactly one, and we show below that the differences between α and PME become more stark in certain subperiods and subsamples where beta is further away from one.

Comparing α and GPME, we find a huge difference in cross-sectional standard deviation. For α the cross-sectional standard deviation is only one third of the cross-sectional standard deviation of GPME. The reason is that the buyout fund beta of 0.9 is very far from the estimated $\gamma = 3.8$. As illustrated by the simulations, GPME includes a significant amount of noise at the fund level when β is far from γ, as is the case here. Figure 2 further illustrates that the distribution of GPME looks very different from the distribution of α for buyout funds, and even more diffuse than in VC. Figure 3 shows that GPME experienced substantial time-series volatility around both the dot-com boom and bust and the global financial crisis, relative to alpha and PME. Thus, although mean GPME is a good metric for risk-adjusted performance for all buyout funds in aggregate, individual fund GPMEs are very noisy estimates of fund-level performance.

Overall, the results demonstrate clearly that the benchmark portfolio approach typically delivers a less noisy measure of individual fund abnormal performance than the (G)PME.

A. *Two-factor models*

Table V applies the benchmark portfolio approach in a two-factor setting. In this analysis, we include the return of one of the four corner portfolios (small/big and low/high book-to market, i.e., SL, SH, BL, BH) of the six size-value portfolios underlying the Fama-French factors as a second factor in the SDF in addition to the market factor. More precisely, to obtain easily interpretable betas, we express the second factor as a log excess return relative to the market factor.[13] Looking at one the four corner portfolio relative to the market allows

[13] With a second factor, x, the SDF becomes $M_h = \exp(\delta h - \gamma_m r_h^m - \gamma_x r_h^x)$. The benchmark portfolio from equation (16) becomes

$$R_h^b = \exp\left\{ r_h^f + \beta' r_h + \frac{1}{2}\beta' \mathrm{diag}(\Sigma_h) - \frac{1}{2}\beta' \Sigma_h \beta \right\}$$

us to explore the size- and value-dimensions simultaneously with only one factor in addition to the market factor.

To interpret the betas, consider the model with the big-value portfolio excess return, BH-M, for buyout (column 5 in Panel B of Table V). The loadings are $\beta_m = 0.920$ and $\beta_{BH-M} = 0.677$. This tells us that buyout funds' risk profile has some similarity with big-value firms. To see this, suppose that we have a linear factor model setting and we regress the excess return on a big-value portfolio on the excess return on the market, $R_m - R_f$, and a big-high portfolio return in excess of the market, $R_{BH-M} = R_{BH} - R_m$. We would get coefficients of 1.0 on the market factor and 1.0 on the second factor, so that the sum of the total factor exposures yields

$$1.0 \times (R_m - R_f) + 1.0 \times (R_{BH} - R_m) = R_{BH} - R_f, \tag{18}$$

which is exactly what it should be. Empirically, we find that the coefficients β_m and β_{BH-M} are not exactly equal to unity, but the pattern of two positive coefficients far from zero is broadly consistent with buyout funds being close to big-value in their risk profile.

Now consider the BL-M case for buyout (column 4 in Panel B). Here we get $\beta_m = 0.920$ and $\beta_{BL-M} = -0.537$. This is also what we should find if buyout funds' risk profile is close to the risk profile of big-value firms. For a rough calculation, suppose that $R_M \approx \frac{1}{2}R_{BL} + \frac{1}{2}R_{BH}$. In a linear factor model setting, the coefficient in a regression of excess return on a big-value portfolio on $R_m - R_f$ and $R_{BL-M} = R_{BL} - R_m$ would be $\beta_m = 1$ and $\beta_{BL-M} = -1$, because with these loadings, the total factor exposure adds up to

$$\begin{aligned} &1.0 \times (R_m - R_f) - 1.0 \times (R_{BL} - R_m) \\ &= \frac{1}{2}R_{BL} + \frac{1}{2}R_{BH} - R_f - R_{BL} + \frac{1}{2}R_{BL} + \frac{1}{2}R_{BH} \\ &= R_{BH} - R_f. \end{aligned} \tag{19}$$

Empirically, our estimates $\beta_m = 0.920$ and $\beta_{BL-M} = -0.537$ are roughly in line with this

where $r = \left[r_h^m - r_h^f; r_h^x\right]$, $\beta = [\beta_m; \beta_x]$, and Σ_h is the covariance matrix of r_h.

prediction. Unlike in the simple illustrative example here, in reality factor variances differ, the market portfolio also has a little small-cap exposure, weights on BH and BL in the market portfolio are not exactly one half, and buyout funds are not exactly the same as the BH portfolio in terms of risk profile, so the simple illustration here is an oversimplification. But broadly, the findings of positive loading on the second factor in the BH-M case and negative loading in the BL-M case both point to the same conclusion: the risk profile of buyout funds resemble the risk profile of the big-value portfolio.

Looking at the two-factor models with the small-cap factors, we see a similar pattern of a positive loading on the value version, SH-M, and a negative loading on the growth version, SL-M. The cross-sectional standard deviation of estimated alphas for the two-factor model with SL-M is the smallest of all four two-factor model versions. [14] This suggests that, instead of big-value, an even better characterization of the risk of buyout funds is the market portfolio after removing small growth exposure.

Turning to VC funds in Panel A, all the results except for the SH-M model are also sensible: large exposure to the market factor with loadings above 2, with only a small correction (relative to the large market exposure) toward big stocks, and away from small stocks. However, the results for SH-M appear unreliable. The estimation error in this model is much higher than in the other models, as indicated by the large p-value of 0.478 despite a large GPME estimate of 0.670, and the cross-sectional standard deviation of α estimates is far higher for the two-factor model with SH-M compared to the other factor models. This is likely the consequence of the beta estimation starting with an unreliable estimate of the mean alpha, due to an unreliable GPME estimate.[15]

The factor model that, for VC, produces the smallest cross-sectional standard deviation

[14] The standard deviation of α is the square root of the objective function Q in equation (17), so the model with the lowest standard deviation is the best fitting model.

[15] We do not have a definite answer as to why we get such imprecise GPME estimates in the SH-M model. We suspect it has to do with the extremely high positive skewness of the estimated SDF in the SH-M case that interacts with the positive skewness of VC fund payoffs (e.g., compare the mean and median IRR or TVPI for VC funds in Table I). With a high coefficient on both the market factor and the SH-M factor, the SDF takes extremely high positive values when both the market factor and the SH-M factor have negative realizations. As a consequence, if a few very large VC fund distributions occur, by chance, at the same time when the SDF realization also happens to be extremely large, the cross-product of SDF and fund payoff will be extremely large, which leads to a very high positive abnormal return estimate, but one that is also very imprecisely estimated.

of α and the largest beta (in absolute value) on the second factor is the one with the BL-M factor. Somewhat surprisingly perhaps, a benchmark portfolio tilted toward large growth stocks (rather than small growth) provides the closest approximation of VC fund payoffs.

Finally, across all different two-factor specifications, we see the same pattern in terms of cross-sectional standard deviations of α and GPME as in the one-factor case. Both in Panel A, for VC, and Panel B, for buyout funds, the dispersion of α is always lower than the dispersion of GPME, consistent with more precisely estimated abnormal performance. Also, as in the one-factor case, the reduction in dispersion with α relative to GPME is much stronger for buyout funds than for VC funds.

B. Subsamples

Our method does not require the assumption that betas among all VC or buyout funds are the same. For example, if one has reason to believe that, say, one subset of VC funds has different systematic risk than another subset of VC funds, one can apply the method to each subset separately. The key tradeoff in deciding how finely to slice the data is between lower bias in individual fund abnormal performance estimates (due to selecting a set with less heterogeneity in true betas) and higher standard errors (due to a smaller number of funds). Therefore, splitting the sample based on a fund characteristic only brings benefits if this characteristic is correlated with the true beta. Forming subgroups randomly, or based on a variable uncorrelated with beta, just leads to a loss in statistical precision without the benefit of reduced bias in performance estimates.

Table VI reports estimates when the performance metrics are estimated on the subsamples of fully liquidated funds, funds raised before and after the year 2000, and funds below and above median size (computed relative to the median fund size in a fund's vintage year).[16] For brevity, the table does not show the SDF parameter estimates or the higher-order moments of the performance distributions.

Table VI shows that the estimated VC betas are fairly constant across the subsamples, ranging from 2.3 to 2.5. The differences between the three performance metrics are therefore

[16]A number of funds are exactly at median size (especially in venture capital) so the numbers of funds in the subsamples of small and large funds are not equal.

comparable to those of the full sample results. Mean α and GPME are between about 0.29 and 0.51 below the mean PME (compared to 0.44 in the full VC sample), and the standard deviations of GPME and PME are substantially higher than the standard deviation of α in most subsamples.[17] The means vary quite substantially across subsamples. Liquidated VC funds have an insignificant mean alpha, which is higher than the full sample mean at least in part because many of these funds were raised earlier in the sample, and Panel B shows that VC funds raised before the year 2000 have positive, but insignificant, mean alpha. Funds raised in 2000 or later, and smaller funds, have significantly negative mean alpha. The variation in alpha across funds also differs by subsample. The standard deviation of α for liquidated funds and those raised before 2000 is about twice as large as that for post-2000 funds. Small funds have somewhat lower variation in alphas than large funds.

In contrast to VC, buyout betas vary more strongly across subsamples. The estimated beta of 0.8 for funds raised after the year 2000 is near the lowest across all subsamples, whereas pre-2000 funds have the highest beta at 1.3. This difference may be due to the fact that leveraged buyout deals were more highly levered in the 1980s and 1990s, differences in the types of industries that underwent buyouts, and differences in buyout strategies. The MSCI-Burgiss coverage is also not very extensive before 1993, so some degree of sample selection is possible. Another potential explanation is that the pre-2000 period contains more liquidated funds. Non-liquidated fund performance is calculated using the most recent NAV, which may not have been updated by GPs (especially prior to the ASC 820/FAS 157 Fair Value accounting standard of 2007) and therefore understate the extent to which performance fluctuates with the benchmark factors. However, this is not likely to be the dominant reason. The beta for the subsample of liquidated funds is close to that of the full sample, not just for buyout but also for VC funds, which are likely more subject to the NAV updating issue than buyout funds. Finally, when we split the sample by fund size, we find that small funds have a beta of 0.8, compared to 1.0 for large funds.

Since the estimated betas for liquidated and large funds are close to one, α and PME

[17] In Panel A and B, for VC funds, the mean α is not exactly equal to the GPME. In these two cases the optimizer could not find a solution for β that yields exact equality. This indicates some degree of misspecification of the log-normal model in these subsamples.

are also close. This is not the case for the post-2000 and small funds, where the mean α is almost 50% higher than the mean PME, due to their low beta. Conversely, the mean alpha of 0.163 for pre-2000 funds is below the mean PME of 0.195, as the estimated beta is above one. For all subsamples, the standard deviations of alpha and PME are close, and fairly constant across subsamples, just somewhat higher for larger and liquidated funds, and somewhat lower for smaller funds. As in the full sample results, the standard deviation of GPME is substantially higher, and the difference to the standard deviation of alpha is larger, for subsamples with lower betas.

Table VII considers three subclasses of venture capital, namely, generalist VCs who invest in companies of all stages, and VCs who invest in early-stage or late-stage startups.[18] All but 80 VC funds are classified into one of these three subclasses. Unlike in the VC subsamples in Table VI, beta estimates show more variation across the subclasses. Generalist VCs are similar to post-2000 funds, with a beta of 2.2, a mean α of -0.230, which is significantly different from zero at the 10% level. Unlike post-2000 funds, the mean PME closer to zero at 0.096, and insignificant. Early-stage VCs have a mean beta of 2.6, while the late-stage beta is 1.7. This drop in beta from early-stage to late-stage funds is consistent with surviving startups turning risky growth opportunities into assets in place as they mature (Berk, Green, and Naik, 1999, and Fisher, Carlson, and Giammarino, 2004, but see Zhang (2005) for a different perspective). As Berk, Green, and Naik (2004) point out, the systematic risk of the early-stage company is higher even if the early-stage risk is purely idiosyncratic (for example, due to technological risk). The mean α of early-stage VCs is -0.192, while late-stage VCs have a positive mean α of 0.019, though neither are significantly different from zero. In contrast, the mean PME of both subclasses is positive, and higher for early-stage than for late-stage funds, at 0.338 and 0.278. Finally, the dispersion in alphas is higher for early-stage VCs compared to late-stage and generalist VCs. This could be due to a higher degree of idiosyncratic risk resulting in more extreme realized outcomes, or larger heterogeneity in skill in early-stage investing.

The variation in estimated betas we find across subsamples and subclasses suggest that

[18]The data do not contain subclass designations for buyout.

there are potential benefits from applying our method separately to these subsets of the data. On the other hand, at least for buyout, the range of betas we estimate, from 0.8 to 1.3, is still well within the range of betas that we consider in the heterogeneous-beta simulations in Table III where the α method under the homogeneous beta assumption still works quite well.

VI. Fund Performance Inference

We introduce two exercises to illustrate the importance of accurate fund-level performance estimates. The first exercise is to replicate common regressions in the literature that have fund performance as the dependent variable. Specifically, we consider regressions that relate performance to fund size. The second exercise considers the persistence of PE performance. All results in this section are based on the full-sample performance estimates reported in Table IV.

A. Performance and fund size

Table VIII reports results of regressions of fund performance on size quartile dummy variables, with the smallest quartile being the omitted category. All regressions include vintage year fixed effects. The regressions in columns 1 and 2 use α as the dependent variable, columns 3 and 4 use GPME, and the final two columns use PME.

The results for venture capital are in Panel A. Column 1 shows that fund alphas are generally increasing in size quartiles with the exception of the third quartile, which has a lower performance than the second quartile. The fourth (largest) quartile has the best performance, which is significantly higher than the first (smallest) quartile, at the 5% level. For GPME in column 3, performance drops from the first to the third quartile, but the result for the largest quartile remains. The standard errors in column 1 are about a third lower than those for GPME in column 3, due to lower degree of noise when using alpha rather than GPME as the dependent variable. All else equal, this increases t-statistics by about 50%. The PME standard errors in column 5 are even higher than their GPME counterparts, except for the fourth quartile (which is slightly lower), and are considerably larger than in the alpha regressions. The increased power of test statistics is an important advantage of using alpha in regressions that include a fund-level performance metric.

Columns 2, 4, and 6 include the natural logarithm of the fund's sequence number as an additional explanatory variable.[19] The coefficient on sequence number is strongly positive and significant. Including fund sequence lowers the size quartile coefficients and renders them insignificant. The effect of size in the earlier regressions appears to capture an unobserved managerial quality effect: Successful VCs survive and get to raise larger funds, and this is captured by the fund's sequence number.

Turning to buyout funds in Panel B, we find marginally significant (at the 10% level) coefficients on the second and third quartile dummies for α and PME. The sequence number coefficients are insignificant for all performance metrics. As in VC, the choice of performance metric has a large effect on standard errors. The standard errors for the alpha regressions are as low as one third the size of those for GPME. Perhaps surprisingly, the adjusted R-squared is higher for the GPME regressions. This is due to the vintage fixed effects soaking up more of the variance in buyout GPMEs compared to alphas. In contrast, the difference between the α and PME regressions is marginal, since the buyout beta is close to one.

Several papers consider the relation between fund size and PME. Kaplan and Schoar (2005) and Harris, Jenkinson, and Kaplan (2014) find similar results for the relation between size and VC performance, though the former paper finds an insignificant coefficient on the fund sequence number, and the latter does not control for sequence. Like us, the literature finds no strong significant relation between buyout fund size and performance (for example, Kaplan and Schoar, 2005, Harris, Jenkinson, and Kaplan, 2014, and Lopez-de-Silanes, Phalippou, and Gottschalg, 2015, Rossi, 2019).

B. Performance persistence

In this section we posit that investors have a preferred (future fund) performance metric in mind, and want to know what measure to use to best predict cross-sectional differences in future fund performance. This could be any variable in principle, but we focus here on past fund performance metrics. To the extent that some GPs have persistent true abnormal

[19] MSCI-Burgiss did not provide fund sequence numbers. We constructed sequence numbers based on GP identifiers and vintage years.

performance that carries over from earlier funds to follow-up funds, a more accurate measure of individual fund abnormal performance should do a better job in revealing this performance persistence.

However, the differences between alternative methods of performance evaluation are more subtle than it might appear at first. In a log-normal framework, the three metrics (α, GPME, and PME) can all be viewed as using the benchmark portfolio R^b, but with different betas: the estimated beta in the case of α, beta equal to the SDF parameter γ in case of the GPME, and beta equal to unity in case of the PME. When the true beta is different from these estimated or implied betas, the abnormal performance produced by the method will have an error component that reflects the estimation error in beta. However, as we show in Appendix D, abnormal performance of funds of the same vintage that have approximately the same lifetimes and the same cash flow profiles during their lifetimes, are affected by approximately the same beta estimation error factor. Hence, the beta estimation error does not affect a cross-sectional performance persistence analysis that predicts future relative performance differences of funds based on relative past performance.

In practice, fund life times are not exactly the same and cash flow profiles different. For this reason, greater accuracy in explicit or implicit beta estimates may result in greater accuracy in cross-sectional performance comparisons. Moreover, and perhaps most importantly, the assumption of jointly log-normally generated fund cash flows and market portfolio returns may not hold. If so, our α measure, which leans heavily on the log-normality assumption, may not work well as a cross-sectional performance comparison criterion. A performance persistence analysis is therefore a useful specification test of the α measure.

We adopt the following procedure. First, for each vintage year between 1985 and 2010 we rank funds by their α metric (we start in 1985 to have a reasonable number of observations for both VC and buyout funds, and we stop in 2010 to allow GPs some time to raise a next fund). For each fund in the top and bottom quartile of its vintage year, we then compute the risk-adjusted performance (measured by α, GPME, and PME) of future funds of the same type (VC or buyout) that were raised by the same GP. Columns 1, 4, and 7 of Table IX show the results for VC in Panel A and buyout in Panel B. The first two rows of each panel show the future performance of GPs with a fund in the top quartile, the second pair of rows show the performance of bottom quartile GPs, and the bottom pair of rows show the differences of top minus bottom. Within each pair of rows, we report the average performance of the next nonoverlapping fund raised,[20] as well as of all future nonoverlapping funds. As a comparison, columns 2, 5, and 8, and 3, 6, and 9 of Table IX repeat the same exercise but using GPME and PME as alternative performance measures to rank funds in the first step of the procedure.

We begin by focusing on VC funds in Panel A and a comparison of sorting on α or PME. For all cases except one (all future funds, with future performance evaluated with PME), sorting on α produces a bigger spread in future performance than sorting on PME does, irrespective of which of the three metrics (α, GPME, PME) is used to assess future fund performance. This is consistent with α benefiting from a much better beta estimate ($\beta = 2.4$) than the $\beta = 1$ that is implicitly hardwired in the PME. At the same time, the differences between α and PME are not huge. This is consistent with the point noted above that much of the effects of beta estimation error affect all cross-sectional units in the same way, which leaves only relatively modest effects of better best estimates on the cross-sectional performance persistence analysis.

The comparison of α and GPME as sorting criterion produces results that are somewhat surprising. Looking across all three blocks of columns in Panel A, we see that sorting on GPME produces a bigger spread in future performance irrespective of which of the three metrics (α, GPME, PME) is used to assess future fund performance. This is likely related to what we found in our simulations in Section IV: While our method produces accurate estimates of beta for a wide range of true betas, the accuracy deteriorates when the true beta is very high. For betas around 3 or higher, we find that our method substantially underestimates the beta. Hence, it could be the case that our estimate for VC funds of 2.4 actually underestimates beta substantially and that the GPME with $\gamma = 3.6$—which can be interpreted as applying the R^b benchmark portfolio with $\beta = \gamma = 3.6$—actually gets closer

[20] Using the next nonoverlapping fund eliminates concerns of mechanical correlation due to overlap in fund life-times (see, for example, Korteweg and Sorensen, 2017, for a discussion), although overlap should be less of a concern for risk-adjusted (as opposed to raw) performance.

to the true beta. This would explain what we find in Panel A.

But the results for buyout funds in Panel B show that it would be a bad choice to rely on the GPME as an individual fund performance metric. GPME only works well at the individual fund level if it happens to be the case that $\beta \approx \gamma$. But for buyout funds, beta is much lower than γ. Accordingly, as Panel B shows, sorting on α dominates sorting on GPME for buyout funds, irrespective of which of the three metrics is used to assess future fund performance. Furthermore, this again illustrates that the large beta estimation error inherent in the GPME assessment of individual buyout fund performance does not appear to induce spurious performance persistence.

Finally, the buyout fund results show that sorting on α also dominates sorting on PME in all cases, irrespective of which of the three metrics is used to assess future fund performance.

The bottom line is that, as a general purpose measure of abnormal performance (i.e., not just in a special case like GPME when $\beta \approx \gamma$ or PME when $\beta \approx 1$), α seems to be the best of the three measures. These results also provide reassurance that the distributional assumptions we used to derive the α measure are not grossly misspecified vis-a-vis the actual data. Otherwise we would not find that α does a good job in uncovering performance persistence irrespective of the metric that is used to evaluate future fund performance.

VII. Conclusion

We introduce a new method to benchmark payoffs of individual private equity funds when only cash flow data are available, but not returns. The method has a clear theoretical foundation in asset pricing theory and, with a one-factor benchmark, it requires estimation of only three parameters. Risk-adjusted payoffs based on this new measure are identical to the GPME measure of Korteweg and Nagel (2016) at the asset class level, but they are less noisy than the GPME at the individual fund level. For this reason, the new measure provides a more accurate assessment of individual fund performance than the PME or its generalized version, although the measures are identical in in the special case where beta happens to be equal to the market price of risk parameter in the SDF. The method assumes log-normality of payoffs and within-asset-class homogeneity of betas, but simulations show that it is robust to reasonably large deviations from these assumptions. The method works well even in smaller data sets.

The empirical results from a large data set of private equity fund cash flows reveal that the variation in estimated risk-adjusted performance across venture capital funds is considerably smaller with our new method than with the PME and GPME. For buyout funds, the cross-sectional standard deviation of abnormal payoffs according to our new measure is only one third of the cross-sectional standard deviation of the GPME. This illustrates clearly the much lower level of noise in the new measure.

We also extend the new measure to a two-factor setting, where the benchmark portfolio combines the market factor with one of the four corner size-value portfolios underling the Fama-French factors. We obtain economically plausible factor exposures. For venture capital funds, we find that a benchmark portfolio tilted toward big-growth stocks best explains individual fund payoffs and hence produces the smallest cross-sectional dispersion abnormal payoffs. For buyout funds, we find the that a tilt away from small-growth works best. The method can be readily extended to include additional benchmark assets. It also has applications to other asset classes such as real estate funds.

Our new measure enhances the statistical power of tests of the determinants of fund-level performance. For example, the lower level of noise in our new measure results in substantially lower standard errors in regressions of fund performance on the size of the fund, and improvements in fund performance predictability.

REFERENCES

Abell, Peter, and Tahir M. Nisar, 2007, Performance effects of venture capital firm networks, *Management Decision* 45, 923–936.

Andonov, Aleksandar, Yael V. Hochberg, and Joshua D. Rauh (2018), Political representation and governance: Evidence from the investment decisions of public pension funds, *Journal of Finance* 73, 2041–2086.

Ang, Andrew, Bingxu Chen, William N. Goetzmann, and Ludovic Phalippou, 2018, Estimating private equity returns from limited partner cash flows, *Journal of Finance* 73, 1751–1783.

Barber, Brad M., and Ayako Yasuda, 2017, The impact of interim performance on private equity fundraising, *Journal of Financial Economics* 124, 172–194.

Berk, Jonathan B., Richard C. Green, and Vasant Naik, 1999, Optimal investment, growth options, and security returns, *Journal of Finance* 54, 1553–1607.

Berk, Jonathan B., Richard C. Green, and Vasant Naik, 2004, Valuation and return dynamics of new ventures, *Review of Financial Studies* 17, 1–35.

Boyer, Brian, Taylor D. Nadauld, Keith P. Vorkink, and Michael S. Weisbach, 2023, Discount rate risk in private equity: Evidence from secondary market transactions, *Journal of Finance*, 78, 835–885.

Braun, Reiner, Tim Jenkinson, and Ingo Stoff, 2017, How persistent is private equity performance? Evidence from deal-level data, *Journal of Financial Economics* 123, 273–291.

Brown, Gregory W., Eric Ghysels, and Oleg Gredil, 2023, Nowcasting net asset values: The case of private equity, *Review of Financial Studies* 36, 945–986.

Brown, Gregory W., Oleg Gredil, and Steven N. Kaplan, 2019, Do private equity funds manipulate reported returns? *Journal of Financial Economics* 132, 267–297.

Campbell, John Y., and Luis M. Viceira, 2002. *Strategic asset allocation: Portfolio choice for long-term investors.* Oxford University Press, Oxford, UK.

Cao, Jerry, and Josh Lerner, 2009, The performance of reverse leveraged buyouts, *Journal of Financial Economics* 91, 139–157.

Carlson, Murray, Adlai Fisher, and Ron Giammarino, 2004, Corporate investment and asset price dynamics: Implications for the cross-section of returns, *Journal of Finance* 59, 2577–2603.

Cavagnaro, Daniel R., Berk A. Sensoy, Yingdi Wang, and Michael S. Weisbach, 2019, Measuring institutional investors' skill at making private equity investments, *Journal of Finance*

74, 3089–3134.

Chakraborty, Indraneel and Michael Ewens, 2018, Managing performance signals through delay: Evidence from venture capital, *Management Science* 64, 2875–2900.

Chung, Ji-Woong, 2012, Performance persistence in private equity funds, Working paper, Chinese University of Hong Kong.

Chung, Ji-Woong, Berk A. Sensoy, Lea Stern, and Michael S. Weisbach, 2012, Pay for performance from future fund flows: The case of private equity, *Review of Financial Studies* 25, 3259–3304.

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

Driessen, Joost, Tse-Chun Lin, and Ludovic Phalippou, 2012, A new method to estimate risk and return of nontraded assets from cash flows: the case of private equity funds, *Journal of Financial and Quantitative Analysis* 47, 511–535.

Gupta, Arpit, and Stijn Van Nieuwerburgh, 2021, Valuing private equity investments strip by strip, *Journal of Finance* 76, 3255–3307.

Gredil, Oleg, Morten Sorensen, and William Waller, 2020, Evaluating private equity performance using stochastic discount factors, Working paper, Tuck School of Business.

Easton, Peter D., Stephannie Larocque, and Jennifer Sustersic Stevens, 2021. Private equity valuation before and after ASC 820, *Journal of Investment Management* 19, 105–135.

Ewens, Michael, Charles M. Jones, and Matthew Rhodes-Kropf, 2013, The price of diversifiable risk in venture capital and private equity, *Review of Financial Studies* 26, 1854–1889.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

Fama, Eugene F., and Kenneth R. French, 2010. Luck versus skill in the cross-section of mutual fund returns, *Journal of Finance* 65, 1915-1947.

Franzoni, Francesco, Eric Nowak, and Ludovic Phalippou, 2012, Private equity performance and liquidity risk, *Journal of Finance* 67, 2341–2373.

Gompers, Paul A., and Josh Lerner, 1997, Risk and reward in private equity investments: The challenge of performance assessment, *Journal of Private Equity* 1, 5–12.

Harris, Robert S., Tim Jenkinson, and Steven N. Kaplan, 2014, Private equity performance: What do we know? *Journal of Finance* 69, 1851–1882.

Harris, Robert S., Tim Jenkinson, Steven N. Kaplan, and Rüdiger Stucke, 2023, Has persistence persisted in private equity? Evidence from buyout and venture capital funds, *Journal of Corporate Finance* 81, 102361.

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

Hochberg, Yael V., Alexander Ljungqvist, and Annette Vissing-Jorgensen, 2014, Informational holdup and performance persistence in venture capital, *Review of Financial Studies* 27, 102–152.

Hochberg, Yael V., and Joshua D. Rauh, 2013, Local overweighting and underperformance: Evidence from limited partner private equity investments, *Review of Financial Studies* 26, 403–451.

Hüther, Niklas, 2023, Do private equity managers raise funds on (sur)real returns? Evidence from deal-level data, *Journal of Financial and Quantitative Analysis* 58, 2959–2992.

Hüther, Niklas, David T. Robinson, Sönke Sievers, and Thomas Hartmann-Wendels, 2020, Paying for performance in private equity: Evidence from venture capital partnerships, *Management Science* 66, 1756–1782.

Jegadeesh, Narasimhan, Roman Kraussl, and Joshua M. Pollet, 2015, Risk and expected returns of private equity investments: Evidence based on market prices, *Review of Financial Studies* 28, 3269–3302.

Jenkinson, Tim, Wayne R. Landsman, Brian R. Rountree, and Kazbi Soonawalla, 2020. Private equity net asset values and future cash flows, *The Accounting Review* 95, 191–210.

Jenkinson, Tim, Miguel Sousa, and Rüdiger Stucke, 2013, How fair are the valuations of private equity funds? Working paper, Oxford University.

Kaplan, Steven N., and Antoinette Schoar, 2005, Private equity performance: Returns, persistence, and capital flows, *Journal of Finance* 60, 1791–1823.

Korteweg, Arthur, 2019, Risk adjustment in private equity returns, *Annual Review of Finan-*

cial Economics 11, 131–152.

Korteweg, Arthur, 2023, Risk and return in private equity. In B. Espen Eckbo, Gordon M. Phillips, and Morten Sorensen (Eds.), *Handbook of the Economics of Corporate Finance, Vol. 1: Private Equity and Entrepreneurial Finance*, 343–418. Elsevier: North Holland.

Korteweg, Arthur, and Stefan Nagel, 2016, Risk-adjusting the returns to venture capital, *Journal of Finance* 71, 1437–1470.

Korteweg, Arthur, Stavros Panageas, and Anand Systla, 2024, Private equity for pension plans? Evaluating private equity performance from an investor's perspective, Working paper, University of Southern California and UCLA.

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

Korteweg, Arthur, and Morten Sorensen, 2017, Skill and luck in private equity performance, *Journal of Financial Economics* 124, 535–562.

Lerner, Josh, Antoinette Schoar, and Wan Wongsunwai, 2007, Smart institutions, foolish choices: The limited partner performance puzzle, *Journal of Finance* 62, 731–764.

Liu, Zhiyang, and Zhiqi Chen, 2014, Venture capital networks and investment performance in China, *Australian Economic Papers* 53, 97–111.

Long, John B., 1990, The numeraire portfolio, *Journal of Financial Economics* 26, 29–69.

Lopez-de-Silanes, Florencio, Ludovic Phalippou, and Oliver Gottschalg, 2015, Giants at the gate: Investment returns and diseconomies of scale in private equity, *Journal of Financial and Quantitative Analysis* 50, 377–411.

Metrick, Andrew, and Ayako Yasuda, 2010, The economics of private equity funds, *Review of Financial Studies* 23, 2303–2341.

Nanda, Ramana, Sampsa Samila, and Olav Sorenson, 2020, The persistent effect of initial success: Evidence from venture capital, *Journal of Financial Economics* 137, 231–248.

Opp, Christian C., 2019, Venture capital and the macroeconomy, *Review of Financial Studies* 32, 4387–4446.

Peng, Liang, 2001, Building a venture capital index, Working paper, University of Colorado, Boulder.

Peters, Ryan, 2018, Volatility and venture capital, Working paper, Tulane University.

Phalippou, Ludovic, 2010, Venture capital funds: Flow-performance relationship and performance persistence, *Journal of Banking and Finance* 34, 568–577.

Phalippou, Ludovic, and Oliver Gottschalg, 2009. The performance of private equity funds, *Review of Financial Studies* 22, 1747–1776.

Robinson, David T., and Berk A. Sensoy, 2013, Do private equity fund managers earn their fees? Compensation, ownership, and cash flow performance, *Review of Financial Studies* 26, 2760–2797.

Robinson, David T., and Berk A. Sensoy, 2016. Cyclicality, performance measurement, and cash flow liquidity in private equity, *Journal of Financial Economics* 122, 521–543.

Rossi, Andrea, 2019, Decreasing returns or mean-reversion of luck? The case of private equity fund growth, Working paper, The Ohio State University.

Sensoy, Berk A., Yingdi Wang, and Michael S. Weisbach, 2014, Limited partner performance and the maturing of the private equity industry, *Journal of Financial Economics* 112, 320–343.

Sorensen, Morten, and Ravi Jagannathan, 2015, The public market equivalent and private equity performance, *Financial Analysts Journal* 71, 43–50.

Stafford, Erik, 2022, Replicating private equity with value investing, homemade leverage, and hold-to-maturity accounting, *Review of Financial Studies* 35, 299–342.

Stucke, Rüdiger, 2011, Updating history, Working paper, University of Oxford.

Woodward, Susan, 2009, Measuring risk for venture capital and private equity portfolios, Working paper, Sand Hill Econometrics.

Zhang, Lu, 2005, The value premium, *Journal of Finance* 60, 67–103.

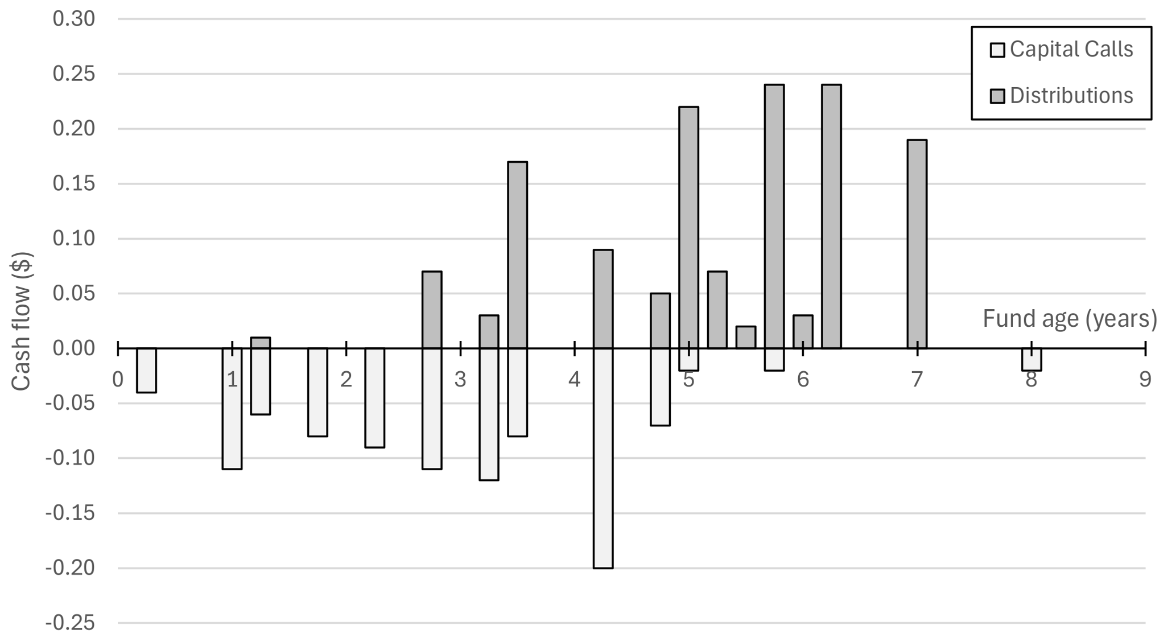


Figure 1. Example cash flow sequence for an individual private equity fund. This figure shows the timing and magnitude of quarterly cash flows to the limited partners of a single private equity fund. The particular fund shown here is a 2012 vintage buyout fund from the MSCI-Burgiss data. All cash flows are net of fees to the general partner, and are scaled to a \$1 commitment. Capital calls are defined as positive numbers in the paper but they are shown as negative cash flows in the graph to emphasize that they are outflows from a limited partner's perspective. At the end of the sample period, which is 9 years after fund inception, the fund reports a remaining net asset value of \$0.40, which is not shown in the graph. The paper treats this final NAV as a distribution that occurs at the end of the sample period.

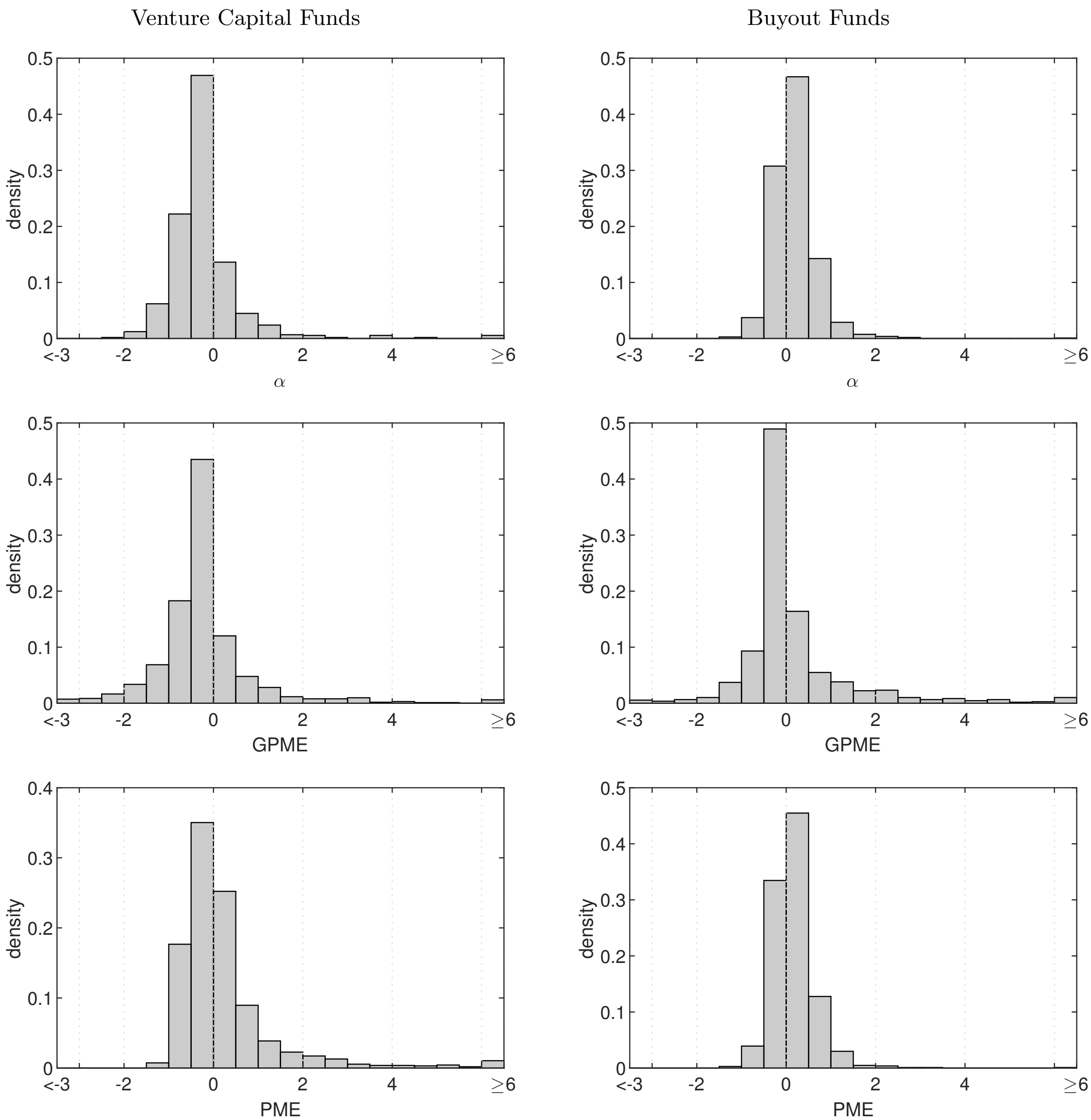
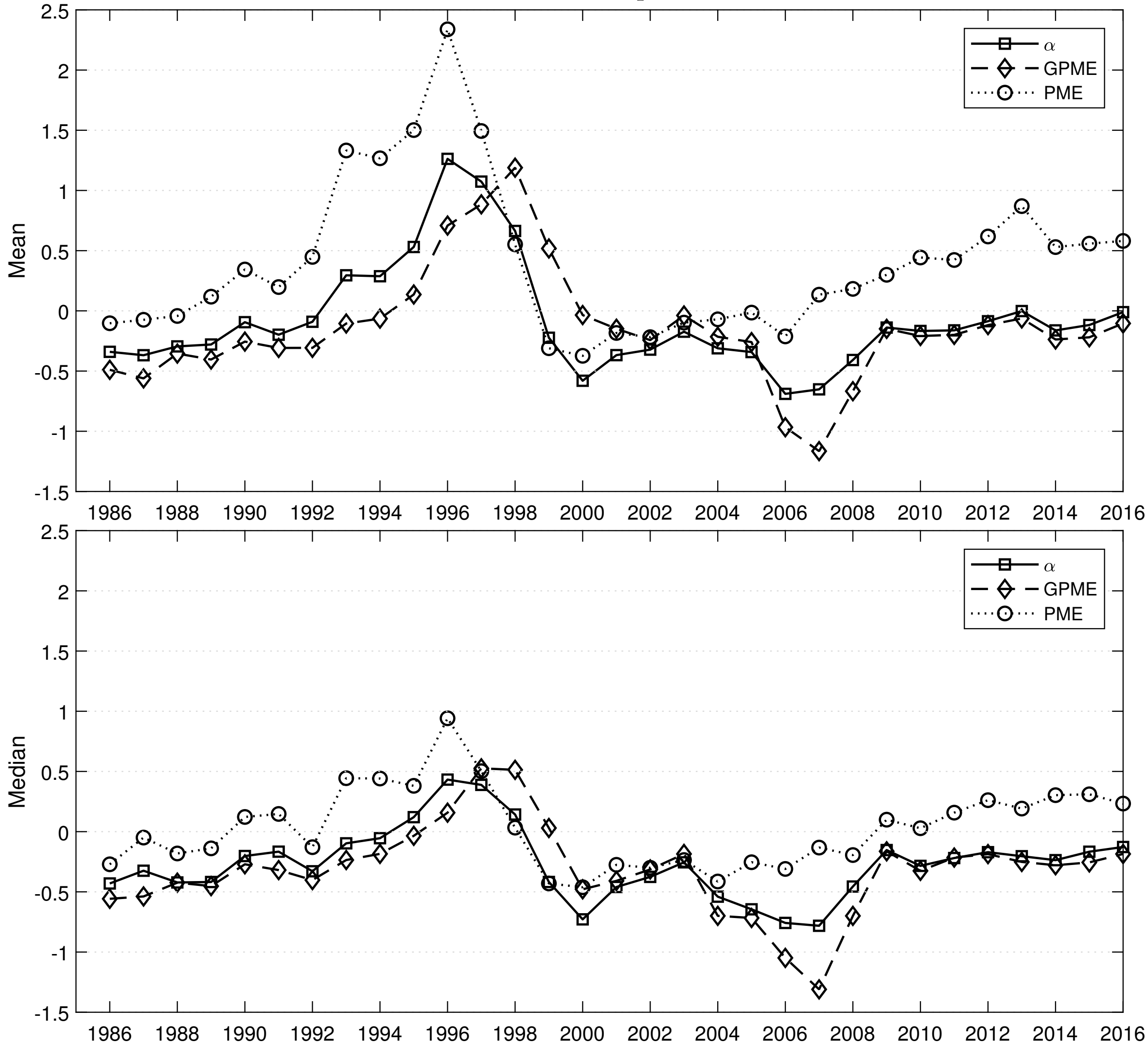


Figure 2. Distribution of performance metrics. This figure shows histograms of fund performance metrics estimated on fund cash flow data from MSCI-Burgiss. The top, middle, and bottom rows show the distribution of fund alphas, GPMEs, and PMEs, respectively. The performance metrics are described in Table IV, using the market return as the sole risk factor in the SDF. The figures in the left and right columns show the performance distribution for venture capital and buyout funds, respectively.

Panel A: Venture Capital Funds



Panel B: Buyout Funds

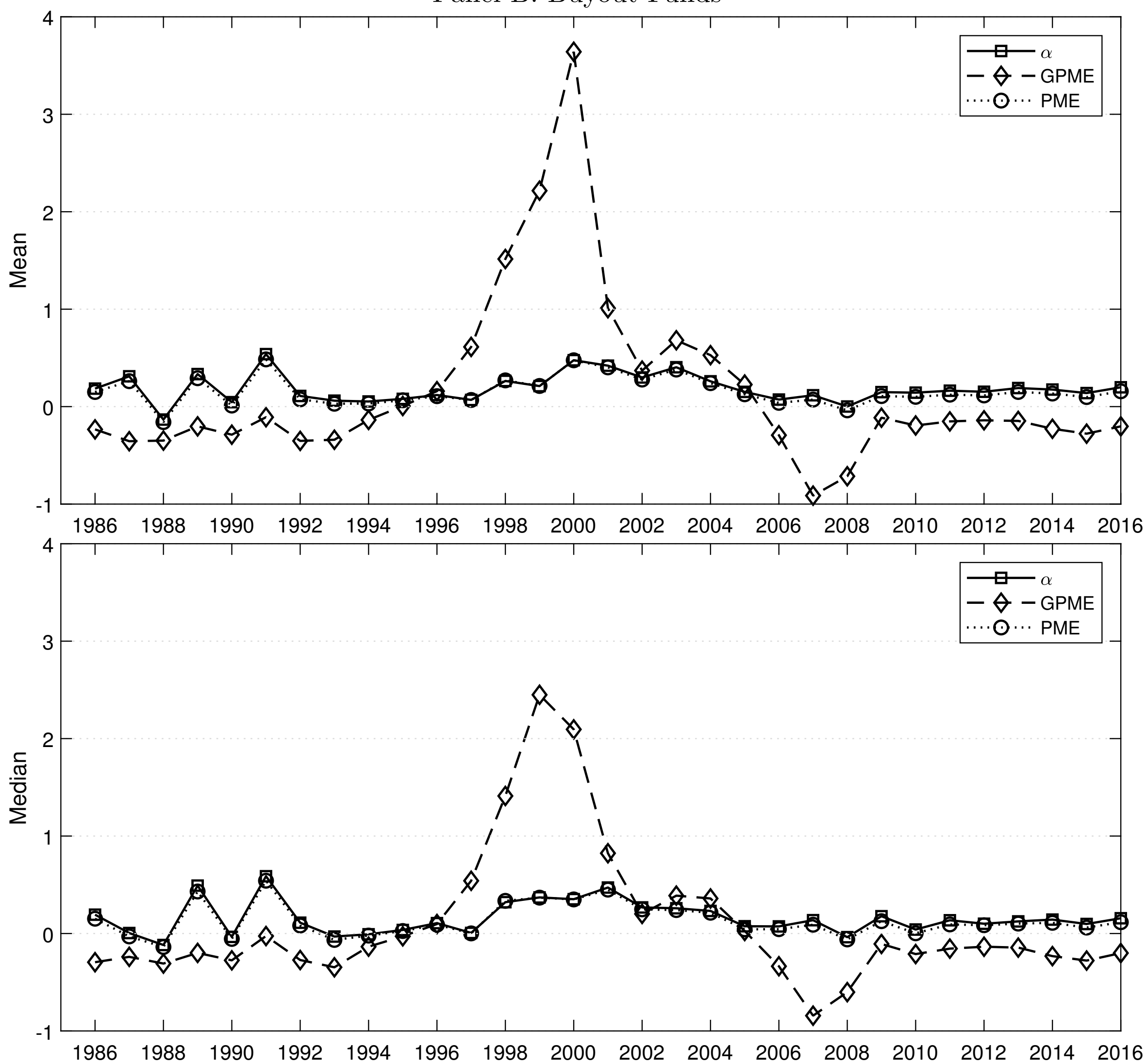


Figure 3. Time series of performance metrics. This figure shows the time series of fund performance metrics for venture capital funds (Panel A) and buyout funds (Panel B) estimated on fund cash flow data from MSCI-Burgiss. The top plot in each panel shows the mean of the alpha, GPME, and PME performance metrics (as described in Table IV, using the market return as the sole risk factor in the SDF) and the bottom plot show the median, by vintage year. Note that the scale on the vertical axis is the same within each panel, but differs across panels.

Table I

Descriptive Statistics

This table reports descriptive statistics of U.S. venture capital and buyout funds from MSCI-Burgiss, with vintages between 1978 and 2016 that have committed capital of at least $5 million in 1990 dollars. Panel A shows statistics across all funds of a given asset class. *Funds* is the number of funds. *Percentage of funds liquidated* is the percentage of funds that are over 10 years old with a final NAV of zero (*100% liquidated*), or the percentage of funds that are over 10 years old with a most recently reported NAV of less than 5% of committed capital (*95% liquidated*), or less than 10% of committed capital (*90% liquidated*). *Firms* is the number of GPs in the sample, and *Funds per firm* is the number of funds raised by a GP. *Fund size* is the total commitment to a fund, in millions of dollars. *Fund effective years* is the time between the first and last observed cash flows of a fund, and *Cash flows per fund* is the number of cash flows observed for a fund (counting multiple cash flows on the same date as one cash flow). Cash flows data runs until the end of the first quarter of 2022. A fund's *IRR* is computed using the final observed NAV of the fund. *TVPI* is total value to paid-in capital, computed as the sum of cash distributions to LPs plus the final NAV divided by the sum of cash takedowns by the fund from LPs. The table also reports fund size weighted averages of the *IRR* and *TVPI*. Panel B reports *IRR* and *TVPI* by vintage year. *N* is the number of funds, and *Wtd* is the size weighted average of the performance measures. N/R stands for "not reported", due to data provider rules that prevent disclosure of performance statistics for sets with fewer than 5 funds.

Panel A: Descriptive Statistics

	Venture Capital Funds			Buyout Funds		
	Mean	Median	St.Dev.	Mean	Median	St.Dev.
Funds	1,630			1,073		
Percentage of funds liquidated:						
100% liquidated	45			40		
95% liquidated	53			51		
90% liquidated	57			54		
Firms	608			436		
Funds per firm	2.68	2.00	2.62	2.46	2.00	1.84
Fund size ($m)	240.70	150.00	289.38	1,116.10	440.97	2,109.27
Fund effective years	13.62	14.16	4.78	12.66	13.01	4.34
Cash flows / fund	38.76	35.00	19.95	55.79	50.00	32.58
IRR (%)	17.16	11.27	40.20	15.88	15.14	16.14
Size-weighted	15.68		32.96	15.63		11.68
TVPI	2.58	1.67	3.34	1.89	1.75	1.04
Size-weighted	2.52		2.89	1.83		0.69

Table I - Continued

Panel B: Performance Statistics by Vintage Year														
	Venture Capital Funds							Buyout Funds						
	N	IRR			TVPI			N	IRR			TVPI		
		Mean	Median	Wtd	Mean	Median	Wtd		Mean	Median	Wtd	Mean	Median	Wtd
<1990	219	11.25	9.53	14.75	2.05	1.74	2.38	46	24.84	15.25	21.98	3.24	2.20	2.91
1990	14	22.62	20.88	28.87	2.85	2.23	3.36	8	17.38	13.74	15.62	2.13	1.75	1.99
1991	7	24.79	23.62	28.17	2.78	2.43	2.83	N/R	N/R	N/R	N/R	N/R	N/R	N/R
1992	18	24.09	14.33	22.09	3.14	1.77	2.67	9	22.82	22.93	35.31	1.97	1.88	2.16
1993	20	49.02	40.51	52.17	5.59	3.20	5.73	9	20.21	18.18	19.25	1.88	1.72	1.87
1994	19	41.88	37.90	59.95	5.42	3.00	9.14	18	16.17	12.77	19.30	1.72	1.46	1.78
1995	29	58.20	30.71	54.31	5.17	2.67	4.93	26	13.57	11.45	12.95	1.61	1.53	1.58
1996	20	87.88	64.07	95.31	5.96	3.07	7.09	17	12.80	9.45	16.43	1.63	1.70	1.74
1997	46	79.03	31.70	80.27	3.60	1.99	3.53	29	5.70	3.32	8.70	1.30	1.21	1.49
1998	57	27.07	6.22	27.53	1.89	1.17	1.97	41	5.72	9.08	3.91	1.45	1.44	1.32
1999	96	-5.12	-6.57	-5.88	0.85	0.72	0.88	33	5.68	9.26	7.17	1.42	1.54	1.46
2000	124	-4.37	-3.12	-1.39	0.95	0.85	1.03	49	14.17	13.50	16.36	1.74	1.60	1.80
2001	64	0.40	2.20	2.93	1.27	1.11	1.48	33	21.88	21.91	24.67	1.84	1.92	1.93
2002	20	0.53	1.43	1.34	1.12	1.08	1.10	19	18.32	18.68	19.93	1.83	1.79	1.97
2003	24	-3.88	2.74	0.79	1.32	1.07	1.66	25	18.74	14.10	23.23	2.03	1.89	2.06
2004	45	1.42	0.34	3.92	1.69	1.01	1.66	43	14.01	12.10	15.13	1.72	1.61	1.83
2005	67	2.80	4.47	7.67	1.84	1.31	2.09	61	9.79	10.18	9.87	1.67	1.61	1.63
2006	87	1.00	4.37	4.09	1.65	1.25	1.83	64	9.26	9.66	7.69	1.66	1.61	1.53
2007	80	10.79	11.03	15.00	2.53	1.68	2.64	65	13.51	13.97	12.42	1.81	1.75	1.73
2008	63	13.52	12.79	14.80	2.71	1.81	2.61	70	12.49	12.50	14.73	1.72	1.65	1.69
2009	29	21.14	21.05	21.17	3.59	2.32	3.37	23	19.55	22.91	20.50	2.19	2.14	2.08
2010	33	20.76	15.66	20.74	3.93	2.49	3.89	31	17.69	16.01	14.14	2.10	1.85	1.88
2011	54	19.79	19.62	22.08	3.67	2.73	3.61	51	18.92	18.99	18.12	2.17	2.05	2.22
2012	60	22.20	20.89	23.56	4.32	2.97	4.61	51	18.49	16.69	19.49	2.00	1.89	2.04
2013	57	21.65	19.72	19.97	4.46	2.64	3.28	49	20.18	19.57	18.06	2.04	1.91	1.96
2014	94	27.45	22.74	29.27	3.29	2.55	3.54	73	18.73	19.05	20.18	1.98	1.87	2.05
2015	106	27.75	27.20	28.74	3.15	2.52	3.26	53	19.47	17.91	21.33	1.92	1.71	1.98
2016	78	34.94	28.39	35.67	3.02	2.38	3.10	74	23.55	23.22	21.89	1.99	1.91	1.90
1990s	326	33.35	9.52	21.56	2.93	1.44	2.30	193	10.77	9.79	9.72	1.58	1.49	1.52
2000s	603	3.40	4.16	6.42	1.78	1.24	1.88	452	13.78	13.19	13.36	1.77	1.69	1.72
≥2010	482	26.07	22.78	26.75	3.57	2.59	3.59	382	19.86	19.46	19.83	2.02	1.89	2.00

Table II

Simulation Results

This table reports summary statistics of the fund-level abnormal performance measures α, GPME and PME, based on a bootstrap sample of funds that takes the capital calls (scaled by each fund's size) of the sample of 2,703 venture capital and private equity funds described in Table I and invests them in the levered public market index, rebalanced monthly to maintain a specific beta. To add a realistic distribution of ourperformance, a random cash flow is added to the first capital call of each fund, drawn from the empirical distribution of fund PMEs (where PME is as defined below). The mean and standard deviation of this true realized fund alpha are 0.242 and 1.187 (irrespective of the choice of beta). The benchmark portfolio used to estimate the α performance metric is

$$R_h^b = \exp(r_h^f + \beta(r_h^m - r_h^f) - \frac{h}{2}\beta(\beta - 1)\sigma^2),$$

for a period from fund inception to h years later. Panel A reports the estimates of the common fund beta, β. The GPME performance measure is computed using an exponential affine stochastic discount factor (SDF)

$$M_h = \exp(\delta h - \gamma r_h^m).$$

PME is a restricted version of the alpha metric that assumes $\beta = 1$ regardless of the true beta, or equivalently, a restricted version of GPME with $\delta = 0$ and $\gamma = 1$.

The columns under the "Raw" header in Panel A report, for each fund performance metric, its mean and its standard deviation (in parentheses) across funds. The columns under the "Excess" header show the mean and standard deviation of excess fund performance, which is defined as the metric minus the true realized fund alpha. Panel B reports the root mean squared error (*RMSE*) relative to true realized fund alphas, and the correlation (*Corr.*) between the metric and true realized alphas.

Panel A: Mean (Standard Deviation)							
β	$\widehat{\beta}$	Raw			Excess		
		α	GPME	PME	α	GPME	PME
0.5	0.490	0.255	0.255	0.143	0.013	0.013	-0.100
		(1.187)	(1.608)	(1.188)	(0.007)	(1.124)	(0.088)
1	1.000	0.242	0.242	0.242	0.000	0.000	0.000
		(1.187)	(1.468)	(1.187)	(0.000)	(0.906)	(0.000)
1.5	1.511	0.212	0.212	0.344	-0.030	-0.030	0.102
		(1.187)	(1.343)	(1.195)	(0.022)	(0.674)	(0.112)
2	1.967	0.174	0.174	0.449	-0.068	-0.068	0.206
		(1.188)	(1.252)	(1.220)	(0.065)	(0.449)	(0.247)
2.5	2.310	0.139	0.139	0.555	-0.103	-0.103	0.313
		(1.194)	(1.206)	(1.267)	(0.130)	(0.266)	(0.404)
3	2.531	0.113	0.113	0.664	-0.130	-0.130	0.422
		(1.208)	(1.203)	(1.339)	(0.213)	(0.213)	(0.578)

Table II - Continued

Panel B: Relation to True Realized Alpha						
β	RMSE			Corr.		
	α	GPME	PME	α	GPME	PME
0.5	0.014	1.124	0.133	1.000	0.716	0.997
1	0.000	0.906	0.000	1.000	0.787	1.000
1.5	0.038	0.674	0.151	1.000	0.865	0.996
2	0.094	0.454	0.322	0.999	0.933	0.979
2.5	0.166	0.285	0.511	0.994	0.975	0.948
3	0.249	0.250	0.716	0.984	0.984	0.902

Table III

Simulations with Heterogeneous Fund Betas

This table shows estimates of α, GPME, and PME for the bootstrap simulation described in Table II, but instead of assuming a common beta for all funds, each fund's true beta is drawn from a normal distribution with mean μ_β and standard deviation 0.25. The true realized fund alpha has a mean and standard deviation (across funds) of 0.242 and 1.187 for all values of μ_β.

Panel A: Mean (Standard Deviation)							
μ_β	$\widehat{\beta}$	Raw			Excess		
		α	GPME	PME	α	GPME	PME
0.5	0.511	0.253	0.253	0.145	0.011	0.011	-0.098
		(1.192)	(1.607)	(1.191)	(0.088)	(1.122)	(0.104)
1	1.017	0.240	0.240	0.244	-0.002	-0.002	0.002
		(1.190)	(1.467)	(1.190)	(0.071)	(0.905)	(0.072)
1.5	1.524	0.211	0.211	0.346	-0.032	-0.032	0.104
		(1.189)	(1.343)	(1.200)	(0.065)	(0.677)	(0.143)
2	1.975	0.174	0.174	0.451	-0.068	-0.068	0.209
		(1.190)	(1.254)	(1.223)	(0.087)	(0.459)	(0.269)
2.5	2.313	0.139	0.139	0.558	-0.103	-0.103	0.316
		(1.196)	(1.210)	(1.274)	(0.143)	(0.287)	(0.422)
3	2.532	0.113	0.113	0.667	-0.129	-0.129	0.425
		(1.209)	(1.208)	(1.348)	(0.221)	(0.241)	(0.596)
Panel B: Relation to True Realized Alpha							
μ_β		RMSE			Corr.		
		α	GPME	PME	α	GPME	PME
0.5		0.089	1.122	0.143	0.997	0.716	0.996
1		0.071	0.905	0.072	0.998	0.787	0.998
1.5		0.073	0.677	0.177	0.999	0.864	0.993
2		0.111	0.464	0.341	0.998	0.931	0.976
2.5		0.176	0.305	0.528	0.993	0.972	0.943
3		0.256	0.273	0.731	0.983	0.980	0.897

Table IV

Performance Estimates

This table reports statistics of the abnormal return measures α, GPME, and the PME across venture capital funds (in the first three columns) and across buyout funds (in the last three columns). The α estimates use a benchmark portfolio, whose return for a period from fund inception to h years later is equation (16) in the main text, reproduced here for reference:

$$R_h^b = \exp\left\{ r_h^f + \beta(r_h^m - r_h^f) - \frac{1}{2}\beta(\beta-1)\sigma_h^2 \right\}.$$

The table also reports the estimates of the common fund beta, β.
The GPME performance measure is computed using an exponential affine stochastic discount factor (SDF)

$$M_h = \exp(\delta h - \gamma r_h^m).$$

The table reports the SDF parameter estimates of δ and γ.
PME is a restricted version of the alpha metric that assumes $\beta = 1$ regardless of the true beta, or equivalently, a restricted version of GPME with $\delta = 0$ and $\gamma = 1$. The reported p-value is for the test that the mean (G)PME is equal to zero. The p-value for mean alpha is identical to that for mean GPME and not reported separately. Standard errors are in parentheses.

	Venture Capital (N=1,630)			Buyout (N=1,073)		
	α	GPME	PME	α	GPME	PME
Mean	-0.184	-0.184	0.258	0.190	0.190	0.161
Percentiles:						
10th	-0.920	-1.193	-0.683	-0.289	-0.740	-0.305
25th	-0.566	-0.632	-0.392	-0.085	-0.334	-0.106
Median	-0.282	-0.293	-0.059	0.130	-0.138	0.098
75th	-0.027	-0.005	0.370	0.379	0.238	0.350
90th	0.469	0.735	1.256	0.719	1.446	0.671
St.Dev.	0.965	1.377	1.407	0.563	1.632	0.553
Skewness	5.638	6.106	5.092	7.790	5.696	7.774
Kurtosis	59.430	79.148	41.365	147.151	66.053	146.155
p-value		0.195	0.000		0.304	0.000
β	2.373			0.908		
SDF Parameters						
δ		0.182	0		0.211	0
		(0.061)			(0.068)	
γ		3.613	1		3.780	1
		(0.690)			(0.648)	

Table V

Performance across Factor Models

This table reports statistics of the abormal return measures α and GPME across factor models specifications. Panel A reports results for venture capital funds, and Panel B for buyout funds. Model (1) reproduces the one-factor results from Table IV, for reference. Models (2) through (5) report results for various two-factor models. In addition to the excess log market return, these models include the log return of one of the corner portfolios of the Fama-French size and book-to-market portfolios in excess of the log market return: small-growth (SL-M), small-value (SH-M), large-growth (BL-M) and large-value (BH-M). The GPME performance measure is thus computed using an exponential affine stochastic discount factor (SDF)

$$M_h = \exp(\delta h - \gamma_m r_h^m - \gamma_x r_h^x),$$

where x is the second factor (SL-M, SH-M, BL-M, or BH-M). The benchmark portfolio has two beta loadings, one on the excess market return (β_m) and one on the return to the factor x (β_x). For GPME, the table shows the p-value for the test that the mean GPME is equal to zero (the p-value for mean alpha is identical and not reported separately). Standard errors are in parentheses.

	α					GPME				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
x =	No x	SL-M	SH-M	BL-M	BH-M	No x	SL-M	SH-M	BL-M	BH-M
Panel A: Venture Capital										
Mean	-0.184	-0.203	0.670	-0.190	-0.186	-0.184	-0.203	0.670	-0.190	-0.186
Percentiles:										
10th	-0.920	-0.958	-0.636	-0.920	-0.936	-1.193	-1.190	-3.354	-1.177	-1.175
Median	-0.282	-0.295	0.237	-0.292	-0.284	-0.293	-0.305	-0.072	-0.298	-0.293
90th	0.469	0.474	2.104	0.464	0.457	0.735	0.701	4.901	0.724	0.727
St.Dev.	0.965	0.966	1.903	0.960	0.960	1.377	1.360	10.342	1.360	1.361
p-value						0.195	0.141	0.478	0.160	0.179
β_m	2.373	2.525	0.652	2.391	2.425					
β_x		-0.183	-0.740	0.481	0.399					
δ						0.182	0.186	0.580	0.179	0.178
						(0.061)	(0.061)	(0.179)	(0.052)	(0.051)
γ_m						3.613	3.692	6.156	3.586	3.559
						(0.690)	(0.668)	(1.398)	(0.676)	(0.633)
γ_x							-0.560	3.679	0.249	-0.148
							(1.477)	(0.938)	(3.086)	(1.191)

Table V - Continued

	α					GPME				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
x =	No x	SL-M	SH-M	BL-M	BH-M	No x	SL-M	SH-M	BL-M	BH-M
					Panel B: Buyout					
Mean	0.190	0.193	0.302	0.207	0.212	0.190	0.193	0.302	0.207	0.212
Percentiles:										
10th	-0.289	-0.299	-0.243	-0.291	-0.298	-0.740	-0.740	-1.014	-0.655	-0.633
Median	0.130	0.121	0.242	0.160	0.163	-0.138	-0.135	-0.023	-0.187	-0.142
90th	0.719	0.740	0.902	0.741	0.757	1.446	1.429	2.527	1.340	1.180
St.Dev.	0.563	0.520	0.607	0.567	0.559	1.632	1.629	2.707	1.730	1.716
p-value						0.304	0.307	0.091	0.318	0.279
β_m	0.908	0.886	0.585	0.920	0.920					
β_x		-1.372	0.122	-0.537	0.677					
δ						0.211	0.209	0.425	0.180	0.172
						(0.068)	(0.067)	(0.087)	(0.051)	(0.052)
γ_m						3.780	3.750	5.006	3.538	3.309
						(0.648)	(0.636)	(0.688)	(0.581)	(0.575)
γ_x							0.164	2.490	3.202	-1.399
							(1.282)	(0.698)	(3.313)	(1.313)

Table VI

Performance Estimates for Subsamples

This table reports statistics of the fund-level abnormal return measures α, GPME, and the PME for subsamples of the MSCI-Burgiss dataset, using the market return as the sole risk factor. Panel A shows performance estimates for the subsample of funds that were fully liquidated by the end of the sample period (the first quarter of 2022). These are funds that are over 10 years old with a final reported net asset value equal to zero. Panel B uses funds raised before the year 2000, while Panel C uses the funds raised in 2000 or later. Panels D and E consider funds with committed capital below or above the median size of funds raised in the same vintage year. Definitions and estimation approach are as described in Table IV.

	Venture capital			Buyout		
	α	GPME	PME	α	GPME	PME
Panel A: Fully Liquidated Funds						
Mean	-0.063	-0.120	0.171	0.257	0.257	0.232
Percentiles:						
10th	-0.998	-0.937	-0.762	-0.361	-0.548	-0.383
Median	-0.323	-0.330	-0.223	0.174	0.039	0.144
90th	0.885	0.713	1.103	0.968	1.335	0.913
St.Dev.	1.338	1.182	1.641	0.769	1.036	0.756
No. of funds	740	740	740	425	425	425
p-value		0.400	0.000		0.027	0.000
β	2.536			0.871		
Panel B: All Funds of Pre-2000 Vintage						
Mean	0.111	0.045	0.363	0.163	0.163	0.195
Percentiles:						
10th	-0.727	-0.733	-0.676	-0.420	-0.446	-0.383
Median	-0.219	-0.260	-0.140	0.017	0.010	0.078
90th	1.013	0.927	1.667	0.880	0.996	0.829
St.Dev.	1.368	1.302	1.849	0.806	0.771	0.852
No. of funds	545	545	545	239	239	239
p-value		0.820	0.000		0.112	0.000
β	2.405			1.304		

Table VI - Continued

	Venture capital			Buyout		
	α	GPME	PME	α	GPME	PME
	Panel C: All Funds of Post-2000 Vintage					
Mean	-0.303	-0.303	0.205	0.223	0.223	0.151
Percentiles:						
10th	-0.968	-1.745	-0.690	-0.226	-0.941	-0.284
Median	-0.301	-0.307	-0.029	0.175	-0.164	0.101
90th	0.195	0.572	1.206	0.741	1.495	0.628
St.Dev.	0.660	1.645	1.118	0.447	2.431	0.431
No. of funds	1,085	1,085	1,085	834	834	834
p-value		0.074	0.000		0.324	0.000
β	2.261			0.814		
	Panel D: Funds below Median Size					
Mean	-0.249	-0.249	0.199	0.204	0.204	0.137
Percentiles:						
10th	-1.012	-1.253	-0.724	-0.382	-0.991	-0.406
Median	-0.344	-0.325	-0.175	0.159	-0.153	0.086
90th	0.484	0.724	1.253	0.828	1.781	0.726
St.Dev.	0.857	1.140	1.407	0.515	2.055	0.495
No. of funds	806	806	806	524	524	524
p-value		0.052	0.007		0.339	0.000
β	2.482			0.806		
	Panel E: Funds at or above Median Size					
Mean	-0.122	-0.122	0.315	0.183	0.183	0.183
Percentiles:						
10th	-0.848	-1.121	-0.598	-0.199	-0.546	-0.199
Median	-0.246	-0.269	0.016	0.101	-0.126	0.101
90th	0.481	0.741	1.272	0.629	1.280	0.629
St.Dev.	1.049	1.569	1.405	0.603	1.255	0.603
No. of funds	824	824	824	549	549	549
p-value		0.446	0.000		0.276	0.000
β	2.284			1.001		

Table VII

Performance Estimates for Venture Capital Subclasses

This table reports statistics of the fund-level abnormal return measures α, GPME, and the PME for subclasses of venture capital funds in the MSCI-Burgiss dataset. Definitions and estimation approach are as described in Table IV.

	Generalist VCs (N = 342)			Early-Stage VCs (N = 1,034)			Late-Stage VCs (N = 174)		
	α	GPME	PME	α	GPME	PME	α	GPME	PME
Mean	-0.230	-0.230	0.096	-0.192	-0.192	0.338	0.019	0.019	0.278
Percentiles:									
10th	-0.852	-1.139	-0.616	-1.054	-1.295	-0.720	-0.565	-0.804	-0.469
Median	-0.290	-0.311	-0.103	-0.307	-0.292	-0.052	-0.097	-0.187	0.064
90th	0.320	0.730	0.718	0.537	0.731	1.603	0.737	0.983	1.247
St.Dev.	0.595	1.017	1.085	1.131	1.516	1.588	0.736	1.089	0.934
p-value		0.068	0.000		0.211	0.000		0.902	0.000
β	2.183			2.554			1.663		

Table VIII

Performance and Fund Characteristics

This table reports regression results using the α, GPME, or PME fund performance metric as the dependent variable. The market return is the sole risk factor in the SDF. Panel A reports results for venture capital funds, and Panel B for buyout funds. Size quartiles are defined by fund type (venture capital or buyout) and decade (pre-1990, 1990s, and 2000 onwards) based on fund sizes measured in 1990 dollars. The smallest size quartile is the omitted category. Some specifications include the natural logarithm of a fund's sequence number, *log(Sequence)*. All regressions include vintage year fixed effects. Heteroscedasticity-consistent standard errors are in parentheses and p-values are in square brackets.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	α	α	GPME	GPME	PME	PME
Panel A: Venture Capital Funds						
Size quartile 2	0.033	-0.006	-0.007	-0.053	0.075	0.016
	(0.055)	(0.057)	(0.068)	(0.071)	(0.097)	(0.099)
	[0.550]	[0.914]	[0.912]	[0.455]	[0.441]	[0.872]
Size quartile 3	0.016	-0.067	-0.051	-0.147	0.082	-0.043
	(0.051)	(0.056)	(0.067)	(0.075)	(0.084)	(0.088)
	[0.750]	[0.233]	[0.449]	[0.049]	[0.330]	[0.625]
Size quartile 4	0.176	0.032	0.251	0.083	0.259	0.041
(largest)	(0.062)	(0.063)	(0.099)	(0.095)	(0.089)	(0.097)
	[0.005]	[0.618]	[0.011]	[0.383]	[0.004]	[0.668]
log(Sequence)		0.131		0.152		0.197
		(0.034)		(0.051)		(0.051)
		[0.000]		[0.003]		[0.000]
Vintage fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.159	0.166	0.118	0.123	0.133	0.140
Number of funds	1,630	1,630	1,630	1,630	1,630	1,630
Panel B: Buyout Funds						
Size quartile 2	0.071	0.072	0.088	0.066	0.072	0.074
	(0.044)	(0.044)	(0.133)	(0.142)	(0.043)	(0.043)
	[0.103]	[0.100]	[0.508]	[0.641]	[0.091]	[0.086]
Size quartile 3	0.102	0.104	0.063	0.015	0.102	0.104
	(0.058)	(0.062)	(0.133)	(0.153)	(0.056)	(0.061)
	[0.077]	[0.097]	[0.634]	[0.924]	[0.072]	[0.088]
Size quartile 4	0.040	0.043	-0.009	-0.092	0.042	0.046
(largest)	(0.041)	(0.049)	(0.146)	(0.195)	(0.040)	(0.048)
	[0.330]	[0.374]	[0.953]	[0.639]	[0.300]	[0.333]
log(Sequence)		-0.004		0.093		-0.005
		(0.025)		(0.085)		(0.025)
		[0.888]		[0.269]		[0.842]
Vintage fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.061	0.061	0.351	0.351	0.067	0.066
Number of funds	1,073	1,073	1,073	1,073	1,073	1,073

Table IX

Performance Persistence of Nonoverlapping Funds

This table reports the performance of future funds of managers (i.e., General Partners) with a fund in the top or bottom quartile in a given vintage year. Panel A reports results for venture capital funds, and Panel B for buyout funds. The first three columns report results from ranking funds in each vintage from 1985 until 2010 by their alpha, GPME, and PME metric, respectively, selecting the top and bottom quartile funds in each vintage, and computing the average alpha of future funds (in the same asset class) of the same managers. The middle and right sets of three columns repeat the exercise but using GPME and PME as the future fund(s) performance metrics. The alpha and GPME metrics use the market return as the sole risk factor in the SDF. The first row only considers the performance of the manager's next fund that is raised at least 10 years after the current fund (such that it does not overlap in time with the current fund). The second row reports performance of all future funds by the same manager that do not overlap with the current fund, or each other. The number of funds over which results are computed are in parentheses. The bottom two rows show the difference between the top and bottom quartile average performance, for the next fund and all future funds, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Future fund(s) metric:	α			GPME			PME		
Current fund sorted on:	α	GPME	PME	α	GPME	PME	α	GPME	PME
Panel A: Venture Capital									
Managers with Top Quartile Funds:									
Next fund	0.43	0.48	0.33	0.64	0.72	0.57	1.06	1.12	0.87
	(134)	(130)	(139)	(134)	(130)	(139)	(134)	(130)	(139)
All future funds	0.17	0.22	0.13	0.27	0.33	0.24	0.79	0.86	0.72
	(160)	(153)	(163)	(160)	(153)	(163)	(160)	(153)	(163)
Managers with Bottom Quartile Funds:									
Next fund	-0.21	-0.25	-0.23	-0.32	-0.34	-0.29	0.32	0.24	0.19
	(44)	(45)	(35)	(44)	(45)	(35)	(44)	(45)	(35)
All future funds	-0.24	-0.26	-0.27	-0.35	-0.34	-0.31	0.29	0.23	0.16
	(48)	(49)	(38)	(48)	(49)	(38)	(48)	(49)	(38)
Difference Top - Bottom:									
Next fund	0.64	0.72	0.56	0.97	1.07	0.85	0.73	0.88	0.69
All future funds	0.41	0.47	0.40	0.62	0.67	0.55	0.50	0.63	0.56

Table IX - Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Future fund(s) metric:		α			GPME			PME	
Current fund sorted on:	α	GPME	PME	α	GPME	PME	α	GPME	PME
				Panel B: Buyout					
Managers with Top Quartile Funds:									
Next fund	0.31	0.20	0.26	0.91	0.25	0.87	0.28	0.17	0.23
	(93)	(82)	(91)	(93)	(82)	(91)	(93)	(82)	(91)
All future funds	0.25	0.18	0.21	0.36	0.01	0.32	0.21	0.15	0.17
	(106)	(92)	(104)	(106)	(92)	(104)	(106)	(92)	(104)
Managers with Bottom Quartile Funds:									
Next fund	0.10	0.29	0.19	-0.18	0.30	-0.07	0.07	0.26	0.16
	(31)	(42)	(30)	(31)	(42)	(30)	(31)	(42)	(30)
All future funds	0.06	0.22	0.15	-0.28	0.01	-0.21	0.03	0.19	0.11
	(35)	(47)	(35)	(35)	(47)	(35)	(35)	(47)	(35)
Difference Top - Bottom:									
Next fund	0.20	-0.09	0.07	1.09	-0.04	0.94	0.21	-0.09	0.07
All future funds	0.19	-0.04	0.06	0.64	0.00	0.53	0.19	-0.04	0.06

Appendix A: Proofs

Proof of equation (7)**.** The left-hand side of equation (7), with the solutions for δ and γ substituted in, equals

$$E[\exp(\delta h - \gamma r_h^m)X_j] = -r^f h + E[x_j] + \frac{1}{2}\sigma_x^2 - \frac{\mu}{\sigma^2} cov(x_j, r_h^m), \tag{A.1}$$

$$= -r^f h + E[x_j] + \frac{1}{2}\sigma_x^2 - \beta\mu h. \tag{A.2}$$

The right-hand side is

$$E\left[\exp\left\{-r^f h + \frac{h}{2}\beta(\beta-1)\sigma^2 - \beta(r_h^m - r^f h)\right\} X_j\right]$$

$$= -r^f h - \frac{h}{2}\beta\sigma^2 - \beta E[r_h^m - r^f h] + E[x_j] + \frac{1}{2}\sigma_x^2, \tag{A.3}$$

$$= -r^f h + E[x_j] + \frac{1}{2}\sigma_x^2 - \beta\mu h, \tag{A.4}$$

where, for the first equality we used $\beta cov(x_j, r_h^m) = \beta^2 h\sigma^2$. Hence, the left and right-hand sides of equation (7) are equal. ■

Appendix B: Observed versus Bootstrapped Fund Cash Flows

This appendix shows how the cash flows from the bootstrapped funds in Section IV compare to those of real-world private equity funds. The main takeaway is that the bootstrapped funds provide a realistic approximation to actual fund cash flow dynamics.

Figure B.1 shows scatterplots of the average observed cash flows for the funds in the MSCI-Burgiss data (on the horizonal axis) against those of the bootstrapped funds (on the vertical axis) for the first 60 quarters since fund inception (i.e., up to a horizon of 15 years). The left and right columns of plots are for venture capital and buyout funds, respectively. The top row uses the simulation setup of Table II, which assumes all bootstrapped funds have the same true beta, except that there is no added random cash flow to the first quarter of the bootstrapped funds (such that outperformance is zero by construction). This allows for a direct quarter-by-quarter comparison of actual cash flows. The true beta is set to the point estimate for the asset class from Table IV. The bottom row of plots uses simulates funds with heterogeneous true betas, as described in Table III, centered around a mean beta equal to the point estimate for each asset class, and also omitting the additional cash flow in the first quarter.

It is evident that cash flows cluster around the 45-degree line through the origin (the solid black line in the scatterplots). For buyout funds, the actual average cash flows are higher than the bootstrapped ones in certain quarters. This is because the outperformance of buyout cannot be matched by the bootstrapped funds, which have zero outperformance by construction. As can be seen in the time series plots of the average cash flows in Figure B.2, these higher observed cash flows mostly occur after the tenth year of the fund's life.

Appendix C: Additional Simulations

This appendix includes two additional simulation exercises to gauge how well the alpha metric approximates true abnormal fund payoffs: i) in a very large sample that aims to approximate asymptotic results, and; ii) on average across many samples that are of similar size as the observed data. These simulations cannot be done with the bootstrap described in the main text. Instead, we use an alternative approach that allows us to simulate an arbitrary number of data sets, each with an arbitrary number of funds and time periods, and with complete control over the distribution of cash flows. This comes at the expense of cash flow distribution timing that is not as close to the observed actual distributions of funds as in the bootstrap. We simulate under the null hypothesis of zero true alpha.

Specifically, we simulate data sets of N funds spanning T years, with an equal number of funds starting each vintage year. The yearly log market return is drawn from a normal distribution with $\mu = 11\%$, and $\sigma = 15\%$. The log risk-free rate is 2% per year. Given these parameters, an SDF with $\delta = 0.15$ and $\gamma = 4$ exactly prices the risk-free rate and public stocks. These SDF parameters are close to the empirical point estimates in the next section. Funds call \$1 in capital in their vintage year, and they distribute $J = 25$ cash flows at random times, drawn from a uniform distribution over funds' ten year lifetimes. Cash flows are generated as

$$X_j = \frac{1}{J}\exp\left(r^f_{h(j)} + \beta(r^m_{h(j)} - r^f_{h(j)}) - \frac{h(j)}{2}\beta(\beta-1)\sigma^2 + \eta_{h(j)}\right). \qquad \text{(C.1)}$$

This specification implies that the true alpha and GPME are zero for all funds. The idiosyncratic log return, η, is iid normally distributed with standard deviation, ω, of 25% per year, and mean $-\frac{h}{2}\omega^2$. We allow the fund-level r_h to be cross-sectionally correlated with a correlation coefficient of 0.1.

Using these assumptions, the simulated funds have realistic return properties. Consider the case true $\beta = 1$, which which roughly corresponds to buyout funds in actual data. In the simulations, we have a mean TVPI of 2.06 and a median of 1.60, which indicates a moderate degree of positive skewness, not too different from the data where Table I in the main text shows a mean TVPI of 1.89 and a median of 1.75 for buyout. Standard deviations are 1.66 in the simulations compared with 1.04 for buyout in actual data. In the case of $\beta = 2.5$, which roughly corresponds to VC funds in actual data, our simulations produce mean TVPI of 5.78 and median of 2.88, compared with the actual data mean of 2.58 and median of 1.67. Standard deviations are 11.04 in the simulations data compared with 3.34 in the actual data. So in the high-beta cases, our simulations, even though they generate cash flows from a log-normal distribution, actually produce a bit too much volatility and positive skewness. Compounding over long horizons can generate substantial skewness and volatility in these simulations.

Large sample results

Table C.I Panel A reports the simulation results for one very large data set that spans $T = 1,000,000$ years and has one fund per vintage year, for a total of $N = 1,000,000$ funds. Although this is implausibly large, it allows us to approximate the asymptotic properties of the α and (G)PME performance metrics. Each row shows results for a different value of the true common fund beta, ranging from 0.5 to 3. For each fund we compute its true realized alpha as $\sum_{j=1}^{J} z_j \exp\left(\eta_{h(j)}\right)$. These true realized alphas are the benchmark for estimated performance metrics, since they are the best estimates of fund-level outperformance one can get if the true beta is known. For all values of beta we use the same draws for the market and idiosyncratic returns, so true realized alphas are by design the same across betas, with a mean of -0.001 (very close to the true alpha of zero) and a standard deviation across funds of 0.571.

Comparing our estimates of α and GPME, a number of results jump out. First, $\bar{\alpha}$ (i.e., average GPME and, mechanically, average α) is a very good estimate of the true alpha, ranging between -0.006 to +0.012 across the different values of beta. Second, the cross-sectional variation in estimated α is close to the variation in true realized alphas for all values of beta, ranging between 0.567 and 0.581. In contrast, the cross-sectional variation in GPMEs is much higher for most betas, and never lower than the variation in estimated α. Third, the root mean squared error (RMSE) of estimated α, measured relative to true realized alphas, is small, ranging from 0.002 to 0.035. This indicates that estimation error in beta and σ^2 contributes very little to the wedge between estimated alphas and the true alpha of zero. In comparison, the RMSE of GPME is very high, with values between 0.374 and 2.409. Fourth, the correlation of estimated α with true realized alphas is essentially one. This is much higher than the correlation between GPME and true realized alphas, which is as low as 0.229 and no higher than 0.841. These results reveal that α is a more accurate measure of fund-level realized abnormal performance than GPME.

As in the bootstrap simulations in the main analysis, when the true $\beta = 1$, the PME exactly equals the true realized alpha for all funds. Fund alphas are very close, but differ slightly because the estimated beta is 1.006, not exactly one. For $\beta \neq 1$, PME performs poorly both as a measure of aggregate outperformance and as a measure of individual fund outperformance. For example, if the true β equals two, the average PME is 0.534, its standard deviation is 1.171, its RMSE is 0.942, and its correlation with true realized alphas is 0.819. Generally, for $\beta > 1$, PME overestimates excess returns because it assumes a discount rate that is too low, and vice versa for $\beta < 1$.

In Panel B, we simulate data with fund betas drawn randomly from a normal distribution with the means ranging from 0.5 to 3 and a standard deviation of 0.25 (so that 95% of fund betas are within 0.5 of the mean). We use the same draws for the market and idiosyncratic returns, so true realized alphas are the same as in Panel A. We find that $\bar{\alpha}$ is not affected by the heterogeneity in betas. This is not surprising, since SDF pricing does not impose any assumption on betas, so the expected GPME (and expected α) should remain unchanged. One important difference is that while PME performed extremely well when beta equals one for all funds, when the average beta is one but there is cross-sectional variation in fund betas, PME performs comparably to α in terms of RMSE and correlation with true realized alphas. When average beta is not one, PME performs much worse than α, as before.

Panels C and D assess the effect of deviations from joint lognormality of market and PE returns. Material violations of this assumption are likely to come from the idiosyncratic PE returns (rather than from market returns). For example, a large fraction of a venture capital fund's investments fail, and most of its return is generated by one or a few "home run" winners. As a result, idiosyncratic fund returns are right-skewed, resembling returns to a portfolio of long call options. The lognormal distribution, though right-skewed by nature, may not fully capture the thick right tail of the empirical distribution. We consider two scenarios that simulate the idiosyncratic log fund return, η, as a mixture of two components. In the first scenario, there is a 10% chance that the fund's cash flow in a given year is magnified by 50%. In the second, the fund's payout in any year is 100% higher with 1% probability. When there is no such "home run" payout, η is drawn from a normal with mean and variance chosen such that the *unconditional* mean and variance of $\exp(\eta)$ are equal to those of the lognormal distribution in Panel A. Relative to the base case, these two scenarios materially increase the skewness and kurtosis of fund IRRs while preserving the first two unconditional moments. We find that these deviations of lognormality have little impact on the properties of α estimates and GPME. The means and variances of estimated α and (G)PME, as well as their RMSEs and correlations with true realized alphas, are very close to Panel A, for all values of beta.

Small sample results

Table C.II shows results from simulated data sets of $N = 1,200$ funds over $T = 30$ years, which is a representative sample size for data sets observed in practice. To assess the small sample properties of the performance metrics, we repeat the simulation 50,000 times and report averages of the statistics across data sets. The mean realized alpha is -0.0003, with a standard deviation of 0.551, on average across data sets.

Panel A shows estimates from the baseline case that assumes a common true fund beta. Because parameter estimates are noisier in smaller data sets, α and GPME are less accurate compared to the large data set. Still, for betas of two or lower, shows that $\bar{\alpha}$ continues to work quite well as a measure of aggregate outperformance, ranging between -0.042 and +0.006. For these values of beta, the estimated α's are close to the true realized fund alphas, with variances ranging from 0.523 to 0.552, RMSEs between 0.090 and 0.126, and correlations of 0.983 and higher. When beta reaches three, $\bar{\alpha}$ clearly diverges from the mean realized alpha, and estimated fund alphas substantially understate the cross-sectional variation in realized alphas, though their correlation is still close to one. The PME metric does not suffer much performance loss compared to the large data set (and does better in many cases), as no parameters need to be estimated. However, PME performed poorly with large amounts of data, and continues to do so in smaller data sets. As before, only when beta is truly equal to one does PME perform well, as it coincides with realized alphas in this special case. However, even in this smaller data set the difference with the α metric remains relatively small. For all other values of beta, the α metric clearly captures both aggregate outperformance and realized fund betas better than PME.

Panel B shows that allowing for heterogeneity in fund betas does not affect the estimated $\bar{\alpha}$, as in the large sample. This is somewhat surprising in the smaller data sets, as the individual fund α and GPME estimates are noisier, which shows up as higher cross-sectional variation, higher RMSEs, and a lower correlation with true realized alphas compared to the case of a common fund beta. Still, the differences with Panel A are small, and we conclude that a modest amount of cross-sectional variation in betas is harmless for α estimates in both large and realistic size data sets.

Results for deviations from lognormality in Panels C and D are similar to the large data set, except that if anything, estimated α is closer to true realized alphas compared to the lognormal simulations of Panel A. However, the true realized alphas are no longer the same as in the previous simulations, due to the new sampling procedure for idiosyncratic returns. Nevertheless, the main takeaway from Panels C and D is clear: departures from lognormality do not appear to have a detrimental effect on the properties of α as a measure of fund performance.

Appendix D: Performance persistence analysis

In this section, we investigate the consequences of beta estimation error on the cross-sectional performance persistence analysis. To provide intuition, we consider a simplified setting with zero log risk-free rate, funds with a single capital call of $1, and a single distribution in one period ahead, and we assume that all funds have identical β. The funds are from the same vintage, with lifetimes that are perfectly aligned, so their capital calls occur at the same time, and the distribution as well. The payoff for fund j is drawn from a log-normal distribution

$$X_j = \exp\left\{a_j + \beta r_m - \frac{1}{2}\beta(\beta-1)\sigma^2 + \varepsilon_j\right\}, \tag{C.2}$$

where a_j is the abnormal return, ε_j is idiosyncratic noise, and r_m is the log market return with variance σ^2 and $\mu = \log E[\exp(r_m)]$. The benchmark portfolio return based on estimated $\hat{\beta}$ is

$$\hat{R}^b = \exp\left\{\hat{\beta} r_m - \frac{1}{2}\hat{\beta}(\hat{\beta}-1)\sigma^2\right\}. \tag{C.3}$$

Then the measured abnormal performance is

$$\hat{\alpha_j} = \frac{X_j}{\hat{R}^b} - 1 = \exp(a_j)\exp(\eta)\exp(\varepsilon_j) - 1, \tag{C.4}$$

where the terms related to beta and beta estimation error are collected in

$$\eta = (\beta - \hat{\beta})r_m - \frac{1}{2}\beta(\beta-1)\sigma^2 + \frac{1}{2}\hat{\beta}(\hat{\beta}-1)\sigma^2. \tag{C.5}$$

Now note that η is constant in the cross-section of funds. Hence, while the beta estimation error may distort the estimate of the *level* of abnormal performance (by scaling it by a factor $\exp(\eta) \neq 1$), it does not change the cross-sectional *ranking* of abnormal performance. For the same reason, it does not induce spurious *cross-sectional* correlation between past and future abnormal performance. To see this, consider abnormal performance of two funds that exist each in one of two periods, 1 and 2:

$$\hat{\alpha}_{j,1} = \exp(a_{j,1}) \exp(\eta_1) \exp(\varepsilon_{j,1}) - 1, \tag{C.6}$$

$$\hat{\alpha}_{j,2} = \exp(a_{j,2}) \exp(\eta_2) \exp(\varepsilon_{j,2}) - 1. \tag{C.7}$$

Here, due to the common beta estimation error, η_1 will typically be positively correlated with η_2. However, this does not affect the cross-sectional correlation between past and future abnormal performance. If $\exp(a_{j,2})$ is uncorrelated with $\exp(a_{j,1})$, i.e., true performance persistence is zero, the cross-sectional correlation between $\hat{\alpha}_{j,1}$ and $\hat{\alpha}_{j,2}$ is zero as well and binning funds by $\hat{\alpha}_{j,1}$ will produce, on average, zero difference in $\hat{\alpha}_{j,2}$.

In our empirical analysis, we construct the performance rankings in the persistence analysis by looking at funds of the same vintage, hence the beta estimation error component $\exp(\eta_1)$ is approximately the same for all funds, as in the analysis above. This means that this beta estimation error component does not alter cross-sectional ranking and it does not give rise to spurious performance persistence—as long as the performance persistence analysis focuses on cross-sectional comparisons between funds.

The above analysis assumes perfectly aligned time periods for past and future fund life-times of funds in the same vintage. What happens in the case of time periods within the same vintage that are not perfectly aligned? In this case the η terms will not be cross-sectionally constant and hence add noise that to some extent distorts cross-sectional performance rankings. To illustrate, suppose that some funds $j = 1, ..., J$ cover period 1a, which partly overlaps with the life-time 1b of other funds $k = 1, ..., K$ of the same vintage. Then

$$\begin{aligned} \hat{\alpha}_{j,1} &= \exp(a_{j,1}) \exp(\eta_{1a}) \exp(\varepsilon_{j,1}) - 1, \\ \hat{\alpha}_{k,1} &= \exp(a_{k,1}) \exp(\eta_{1b}) \exp(\varepsilon_{k,1}) - 1. \end{aligned} \tag{C.8}$$

Now if 1a ends earlier than 1b, then these early funds may have follow-up funds in time period 2a that, while non-overlapping with 1a, overlap with 1b. So, for the follow-up funds we then have

$$\hat{\alpha}_{j,2} = \exp(a_{j,2}) \exp(\eta_{2a}) \exp(\varepsilon_{j,2}) - 1,$$
$$\hat{\alpha}_{k,2} = \exp(a_{k,2}) \exp(\eta_{2b}) \exp(\varepsilon_{k,2}) - 1. \tag{C.9}$$

where 1b will partly overlap with 2a. Because of this overlap, it is likely, that $\exp(\eta_{1b})$ and $\exp(\eta_{2a})$ will have some positive correlation. There are then two effects of beta estimation error on a cross-sectional performance persistence analysis. Both induce a bias in the estimated cross-sectional correlation of past and future performance:

(i) The first one is due to the overlap. Consider again the case of zero correlation between past true abnormal performance $\exp(a_{j,1})$, $\exp(a_{k,1})$ and future true abnormal performance $\exp(a_{j,2})$, $\exp(a_{k,2})$. If it happens that $\exp(\eta_{1b}) > \exp(\eta_{1a})$, then it will tend to be the case, due to the overlap, that $\exp(\eta_{2a}) > \exp(\eta_{2b})$. Hence, it will tend to be the case that $\hat{\alpha}_{k,1} > \hat{\alpha}_{j,1}$ and $\hat{\alpha}_{k,2} < \hat{\alpha}_{j,2}$, i.e., negative cross-sectional correlation even though the correlation of true past and future abnormal performance is zero.

(ii) The second one is an attenuation bias. If $\exp(\eta_{1a}) \neq \exp(\eta_{1b})$, this adds noise that distorts the performance ranking to some degree.

For these two reasons, there is a benefit from getting the estimate of β as close to the true β as possible. Our proposed α measure may therefore have an advantage in uncovering performance persistence if it gets the estimated beta closer to the truth than the PME (which implicitly assumes $\beta = 1$) or the GPME (which works at the individual fund level only if $\beta = \gamma$).

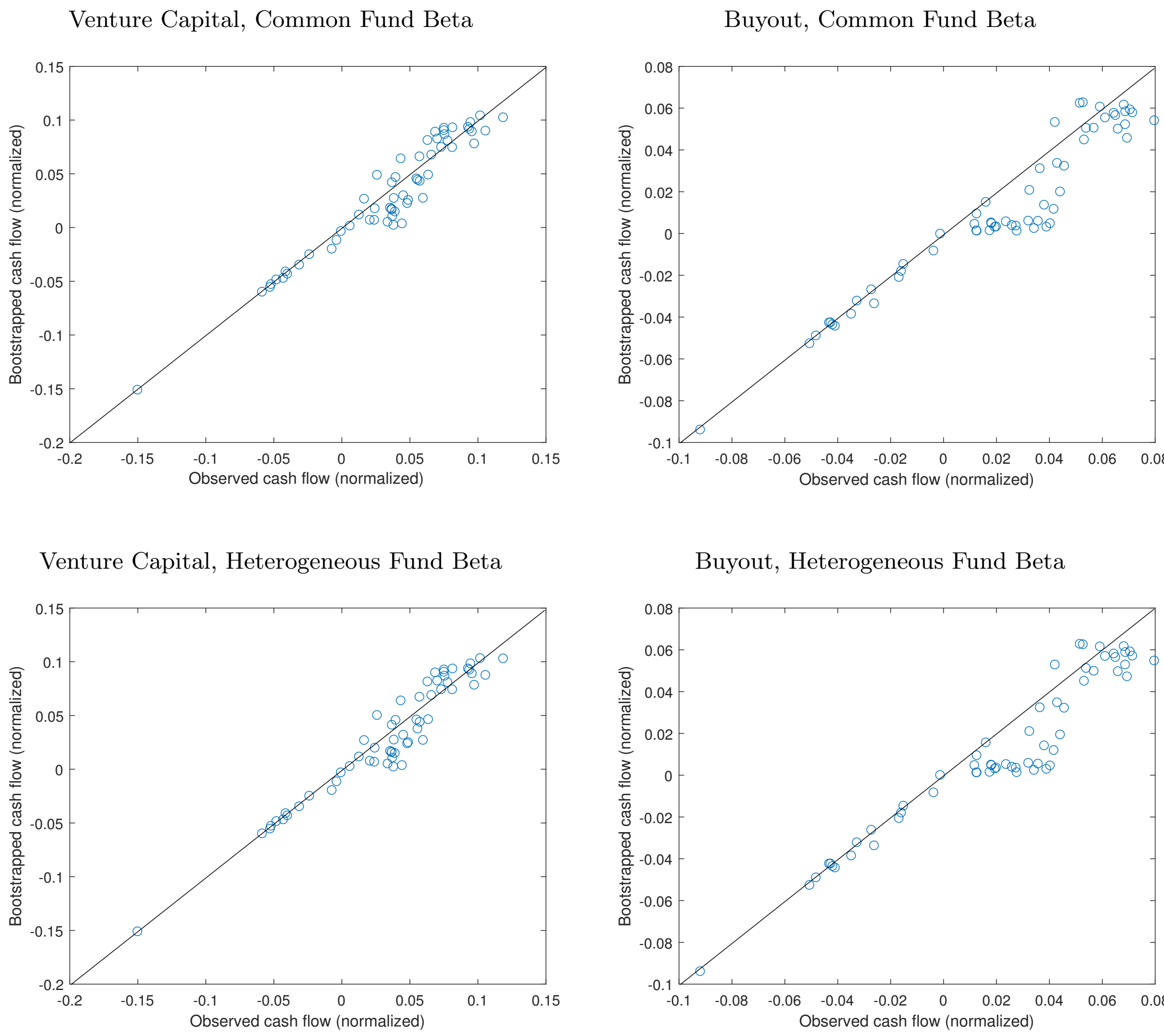


Figure B.1. Scatterplots of observed versus bootstrapped average fund cash flows. This figure plots the average quarterly fund cash flows since inception for the MSCI-Burgiss data (see Section III) on the horizontal axis, against those of the bootstrapped funds (Section IV) on the vertical axis. Cash flows are normalized by fund size. Each data point represents a quarter, measured since fund inception, up to a horizon of 15 years. Negative numbers are capital calls and positive numbers are distributions. The solid black line is the 45-degree line through the origin. The left and right columns of plots are for venture capital and buyout funds, respectively. The top row simulations assume that all bootstrapped funds have the same true beta (as in Table II). The bottom row allows for heterogeneous betas across bootstrapped funds (as in Table III). Unlike the main paper, the bootstrapped funds have no additional cash flow in the first quarter, such that true outperformance is zero by construction. The true (mean) beta for the bootstrapped funds are the point estimates from Table IV.

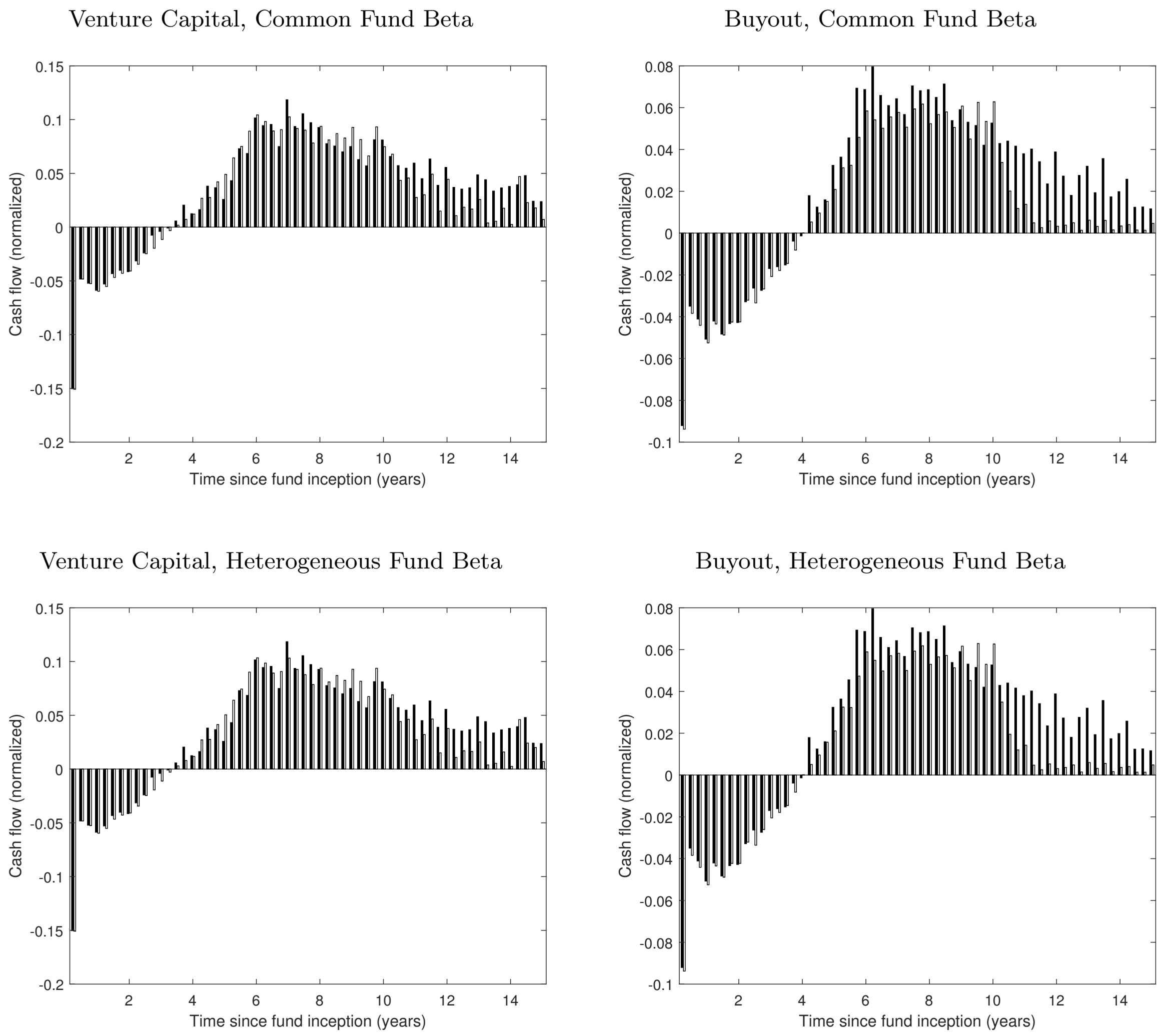


Figure B.2. Time series of observed and bootstrapped average fund cash flows. This figure shows the time series of average quarterly fund cash flow since inception for the MSCI-Burgiss data (see Section III) and the bootstrapped funds (Section IV). The horizontal axis is the time since fund inception in years, up to a horizon of 15 years. The black bars represent the observed funds and the white bars are for the bootstrapped funds. Cash flows are normalized by fund size. Negative numbers are capital calls and positive numbers are distributions to LPs. The left and right columns of plots are for venture capital and buyout funds, respectively. The top row simulations assume that all bootstrapped funds have the same true beta (as in Table II). The bottom row allows for heterogeneous betas across bootstrapped funds (as in Table III). Unlike the main paper, the bootstrapped funds have no additional cash flow in the first quarter, such that true outperformance is zero by construction. The true (mean) beta for the bootstrapped funds are the point estimates from Table IV.

Table C.I

Large Sample Simulation Results

This table reports summary statistics of the fund-level abnormal return measures α, GPME and PME, based on one simulated data set of $T = 1,000,000$ vintages, with one fund per vintage. Funds call \$1 in capital in their vintage year, and distribute 25 cash flows at times drawn randomly from a uniform distribution over their ten year lifetimes. The underlying fund returns, on which cash flows are based, are generated based on an assumed true market β, with an idiosyncratic volatility of 25% per year. The log risk-free rate is 2% per year, and the annual log market return is drawn from a normal distribution with $\mu = 11\%$ and $\sigma = 15\%$. The benchmark portfolio return from fund inception to h years later, which is used to estimate α, is

$$R_h^b = \exp(r_h^f + \beta(r_h^m - r_h^f) - \frac{h}{2}\beta(\beta-1)\sigma^2),$$

The GPME performance measure is computed using an exponential affine stochastic discount factor (SDF)

$$M_h = \exp(\delta h - \gamma r_h^m).$$

PME is a restricted version of the α metric that assumes $\beta = 1$ regardless of the true beta, or equivalently, a restricted version of GPME with $\delta = 0$ and $\gamma = 1$. The simulations in Panel A assume that all funds have the same true beta, shown in the first column. The estimated $\widehat{\beta}$ is in the second column. Panel B allows for heterogeneous fund betas generated from a normal distribution whose mean is in the first column, and a 0.25 standard deviation. Panel C uses the same setup as Panel A, except that there is a 10% probability that idiosyncratic PE returns are 50% higher in any given year. In Panel D, there is a 1% chance of a 100% higher payout. The mixture distribution of the idiosyncratic return in Panels C and D has the same mean and variance as the lognormal in Panel A. For each performance metric, the table reports its mean and standard deviation (in parentheses) across funds, its root mean squared error (RMSE) relative to true realized fund alphas, and the correlation (Corr.) between the metric and true realized alphas (in square brackets).

β	$\widehat{\beta}$	α		GPME		PME	
		Mean (St.Dev)	RMSE [Corr.]	Mean (St.Dev)	RMSE [Corr.]	Mean (St.Dev)	RMSE [Corr.]
Panel A: Common Fund Beta, Lognormal Returns							
0.5	0.511	-0.006	0.008	-0.006	2.409	-0.180	0.265
		(0.567)	[1.000]	(2.475)	[0.229]	(0.471)	[0.948]
1	1.006	-0.003	0.004	-0.003	1.645	-0.001	0.000
		(0.569)	[1.000]	(1.741)	[0.327]	(0.571)	[1.000]
1.5	1.497	0.000	0.002	0.000	1.183	0.231	0.379
		(0.571)	[1.000]	(1.314)	[0.436]	(0.778)	[0.947]
2	1.983	0.004	0.009	0.004	0.854	0.534	0.942
		(0.574)	[1.000]	(1.030)	[0.558]	(1.171)	[0.819]
2.5	2.460	0.008	0.019	0.008	0.595	0.933	1.819
		(0.577)	[1.000]	(0.829)	[0.696]	(1.884)	[0.668]
3	2.917	0.012	0.035	0.012	0.374	1.460	3.251
		(0.581)	[0.999]	(0.690)	[0.841]	(3.163)	[0.526]

Table C.I - Continued

Panel B: Heterogeneous Betas							
0.5	0.528	-0.006	0.008	-0.006	2.374	-0.174	0.257
		(0.567)	[1.000]	(2.441)	[0.232]	(0.473)	[0.951]
1	1.023	-0.003	0.004	-0.003	1.626	0.006	0.011
		(0.569)	[1.000]	(1.723)	[0.331]	(0.576)	[1.000]
1.5	1.514	0.001	0.002	0.001	1.170	0.240	0.394
		(0.572)	[1.000]	(1.302)	[0.439]	(0.787)	[0.944]
2	1.999	0.004	0.009	0.004	0.844	0.546	0.966
		(0.574)	[1.000]	(1.021)	[0.563]	(1.189)	[0.814]
2.5	2.476	0.008	0.019	0.008	0.587	0.949	1.857
		(0.577)	[1.000]	(0.823)	[0.701]	(1.917)	[0.663]
3	2.932	0.012	0.036	0.012	0.367	1.481	3.314
		(0.581)	[0.998]	(0.686)	[0.845]	(3.222)	[0.521]
Panel C: Non-Lognormality (10% Probability of 50% Higher Cash Flow)							
0.5	0.503	-0.000	0.002	-0.000	2.865	-0.178	0.264
		(0.569)	[1.000]	(2.922)	[0.197]	(0.471)	[0.948]
1	0.998	0.002	0.001	0.002	1.732	0.001	0.000
		(0.570)	[1.000]	(1.826)	[0.316]	(0.570)	[1.000]
1.5	1.488	0.005	0.007	0.005	1.196	0.233	0.376
		(0.573)	[1.000]	(1.328)	[0.435]	(0.774)	[0.948]
2	1.972	0.009	0.014	0.009	0.856	0.535	0.932
		(0.575)	[1.000]	(1.033)	[0.560]	(1.158)	[0.820]
2.5	2.446	0.013	0.025	0.013	0.594	0.933	1.788
		(0.579)	[0.999]	(0.831)	[0.699]	(1.849)	[0.670]
3	2.897	0.017	0.044	0.017	0.372	1.458	3.172
		(0.582)	[0.998]	(0.691)	[0.844]	(3.076)	[0.528]
Panel D: Non-Lognormality (1% Probability of 100% Higher Cash Flow)							
0.5	0.501	0.001	0.001	0.001	3.030	-0.178	0.267
		(0.584)	[1.000]	(3.083)	[0.186]	(0.480)	[0.949]
1	0.998	0.002	0.002	0.002	1.781	0.002	0.000
		(0.585)	[1.000]	(1.872)	[0.308]	(0.584)	[1.000]
1.5	1.489	0.006	0.007	0.006	1.207	0.234	0.382
		(0.587)	[1.000]	(1.340)	[0.435]	(0.795)	[0.948]
2	1.973	0.009	0.014	0.009	0.858	0.537	0.954
		(0.590)	[1.000]	(1.040)	[0.565]	(1.195)	[0.821]
2.5	2.447	0.013	0.025	0.013	0.595	0.936	1.849
		(0.593)	[0.999]	(0.839)	[0.705]	(1.926)	[0.669]
3	2.898	0.018	0.044	0.018	0.373	1.465	3.328
		(0.597)	[0.998]	(0.703)	[0.848]	(3.254)	[0.524]

Table C.II

Small Sample Simulation Results

This table reports summary statistics of the fund-level abnormal return measures α, GPME and PME, based on 50,000 simulated data sets of $N = 1,200$ funds over $T = 30$ vintages. Fund-level cash flows and performance metrics are as described in Table C.I. The simulations in Panel A assume that all funds have the same true beta, shown in the first column. The estimated $\widehat{\beta}$ is in the second column. Panel B allows for heterogeneous fund betas generated from a normal distribution whose mean is in the first column, and a 0.25 standard deviation. Panel C uses the same setup as Panel A, except that there is a 10% probability that idiosyncratic PE returns are 50% higher in any given year. In Panel D, there is a 1% chance of a 100% higher payout. The mixture distribution of the idiosyncratic return in Panels C and D has the same mean and variance as the lognormal in Panel A. For each performance metric, in each simulated data set, we compute its mean and standard deviation across funds, its root mean squared error (RMSE) relative to true realized fund alphas, and the correlation between the metric and true realized alphas (in square brackets). The table reports the mean of these metrics across the 50,000 data sets.

β	$\widehat{\beta}$	α		GPME		PME	
		Mean (St.Dev)	RMSE [Corr.]	Mean (St.Dev)	RMSE [Corr.]	Mean (St.Dev)	RMSE [Corr.]
Panel A: Baseline							
0.5	0.499	0.006	0.092	0.006	1.644	-0.179	0.251
	(0.191)	(0.552)	[0.995]	(1.717)	[0.313]	(0.447)	[0.966]
1	1.001	0.000	0.090	0.000	1.538	-0.000	0.000
	(0.207)	(0.548)	[0.994]	(1.623)	[0.344]	(0.551)	[1.000]
1.5	1.514	-0.017	0.097	-0.018	1.413	0.230	0.343
	(0.245)	(0.536)	[0.990]	(1.510)	[0.376]	(0.732)	[0.967]
2	2.033	-0.042	0.126	-0.048	1.281	0.531	0.818
	(0.317)	(0.523)	[0.983]	(1.384)	[0.407]	(1.035)	[0.891]
2.5	2.535	-0.068	0.178	-0.087	1.153	0.927	1.486
	(0.407)	(0.511)	[0.974]	(1.253)	[0.436]	(1.518)	[0.804]
3	3.011	-0.093	0.248	-0.133	1.042	1.451	2.438
	(0.508)	(0.502)	[0.962]	(1.125)	[0.462]	(2.264)	[0.722]

Table C.II - Continued

Panel B: Heterogeneous Betas							
0.5	0.518	0.005	0.195	0.005	1.642	-0.174	0.266
	(0.195)	(0.577)	[0.952]	(1.715)	[0.313]	(0.463)	[0.938]
1	1.017	-0.001	0.179	-0.001	1.538	0.006	0.148
	(0.211)	(0.567)	[0.958]	(1.622)	[0.343]	(0.577)	[0.964]
1.5	1.527	-0.019	0.168	-0.020	1.415	0.239	0.412
	(0.249)	(0.551)	[0.962]	(1.510)	[0.374]	(0.772)	[0.930]
2	2.042	-0.043	0.176	-0.049	1.285	0.543	0.888
	(0.321)	(0.535)	[0.962]	(1.386)	[0.405]	(1.094)	[0.860]
2.5	2.539	-0.068	0.212	-0.088	1.159	0.943	1.574
	(0.410)	(0.520)	[0.957]	(1.256)	[0.433]	(1.603)	[0.780]
3	3.011	-0.093	0.271	-0.133	1.050	1.472	2.558
	(0.510)	(0.511)	[0.948]	(1.130)	[0.458]	(2.387)	[0.705]
Panel C: Non-Lognormality (10% Probability of 50% Higher Cash Flow)							
0.5	0.501	0.006	0.058	0.006	1.651	-0.178	0.253
	(0.122)	(0.563)	[0.998]	(1.730)	[0.319]	(0.455)	[0.966]
1	1.002	0.000	0.056	0.000	1.544	0.000	0.000
	(0.134)	(0.559)	[0.998]	(1.636)	[0.349]	(0.561)	[1.000]
1.5	1.516	-0.018	0.068	-0.018	1.420	0.231	0.346
	(0.173)	(0.547)	[0.995]	(1.522)	[0.381]	(0.746)	[0.967]
2	2.036	-0.042	0.109	-0.048	1.287	0.533	0.826
	(0.263)	(0.534)	[0.988]	(1.396)	[0.412]	(1.055)	[0.892]
2.5	2.540	-0.068	0.171	-0.087	1.158	0.930	1.503
	(0.369)	(0.522)	[0.978]	(1.263)	[0.441]	(1.548)	[0.806]
3	3.014	-0.093	0.246	-0.132	1.047	1.455	2.469
	(0.479)	(0.513)	[0.966]	(1.135)	[0.466]	(2.310)	[0.726]
Panel D: Non-Lognormality (1% Probability of 100% Higher Cash Flow)							
0.5	0.503	0.006	0.074	0.006	1.646	-0.178	0.252
	(0.175)	(0.553)	[0.997]	(1.720)	[0.313]	(0.448)	[0.965]
1	1.004	0.000	0.072	0.000	1.539	0.000	0.000
	(0.186)	(0.549)	[0.996]	(1.627)	[0.343]	(0.552)	[1.000]
1.5	1.519	-0.018	0.082	-0.018	1.415	0.231	0.344
	(0.227)	(0.538)	[0.992]	(1.513)	[0.375]	(0.735)	[0.967]
2	2.039	-0.042	0.117	-0.048	1.283	0.533	0.821
	(0.300)	(0.525)	[0.985]	(1.387)	[0.407]	(1.039)	[0.890]
2.5	2.541	-0.068	0.174	-0.087	1.154	0.930	1.492
	(0.396)	(0.513)	[0.975]	(1.256)	[0.436]	(1.524)	[0.802]
3	3.017	-0.093	0.247	-0.132	1.044	1.454	2.448
	(0.498)	(0.504)	[0.964]	(1.128)	[0.461]	(2.273)	[0.720]