Network for Studies on Pensions, Aging and Retirement

Netspar DISCUSSION PAPERS

Arthur Korteweg and Morten Sorensen

Skill and Luck in Private Equity Performance

DP 04/2014-026

Skill and Luck in Private Equity Performance

Arthur Korteweg

Morten Sorensen

April 2014

Abstract

We evaluate the performance of private equity ("PE") funds, using a variance decomposition model to separate skill from luck. We find a large amount of long-term persistence, and skilled PE firms outperform by 7% to 8% annually. But this performance is noisy, with a large amount of luck, so top-quartile performance does not necessarily imply top-quartile skills, making it difficult for investors ("LPs") to identify skilled PE firms. Buyout ("BO") firms show the largest skill differences, implying the greatest long-term persistence. Venture capital ("VC") performance is the most noisy, making good VC firms hardest to identify, and implying the smallest amount of investable persistence.

The authors can be reached at Stanford Graduate School of Business (korteweg@stanford.edu) and Columbia Business School (ms3814@columbia.edu). We are grateful to Peter Cornelius, Paul Pfleiderer, Matt Rhodes-Kropf, Per Stromberg, seminar participants at the University of Alberta, the University of San Diego, the University of Virginia, and participants at the 2013 Spring JOIM conference on Private Equity for helpful comments and feedback.

The persistence and predictability of returns is a central topic in finance. Studies of individual stocks, mutual funds, and hedge funds generally find that returns are unpredictable and that investors cannot consistently outperform the market. An important exception is private equity ("PE"), including venture capital ("VC"), buyout ("BO") and other types of PE firms. A PE firm typically manages a sequence of PE funds, and Kaplan and Schoar [2005] find that the performance of fund number $N-1$ predicts the performance of the subsequent fund N. A natural interpretation is that PE firms differ in their skills and abilities, and that funds that are managed by skilled PE firms consistently outperform. It is puzzling, then, that this outperformance is not competed away by the investors ("LPs") in these funds, e.g., by driving up the fees that more skilled PE firms charge their LPs.

Kaplan and Schoar [2005], and subsequent studies,[1] measure persistence as a positive and statistically significant β coefficient in the regression:

$$y_{i,N} = \alpha + \beta \times y_{i,N-1} + \varepsilon_{i,N}\,, \tag{1}$$

where $y_{i,N}$ is the performance of fund number N managed by PE firm i. This regression is motivated by a cross-sectional intuition. In the cross-section, some funds have better performance, and if these funds follow previous funds that also had better performance, this is evidence of persistence. Formally, however, equation (1) is a time-series AR(1) model, and there is a tension between this cross-sectional intuition and its time-series properties. The AR(1) model does not distinguish skill from luck. If an unskilled PE firm is lucky and its fund outperforms, then the AR(1) model implies that this PE firm is now considered a skilled firm, and its next fund is also expected to outperform. Conversely, a skilled firm with an unlucky fund is immediately considered unskilled. In the limit, after all PE firms have undergone a number of such transitions, the AR(1) model implies that their performance converges to the same limit distribution, with $E[y] = \frac{\alpha}{1-\beta}$.[2] Hence, the AR(1) model is an empirical model of performance persistence that implies no long-term performance

[1]Including Phalippou and Gottschalg [2009], Hochberg, Ljungqvist, and Vissing-Jorgensen [2014], Phalippou [2010], Robinson and Sensoy [2011], Chung [2012], Braun, Jenkinson, and Stoff [2013], and Harris, Jenkinson, Kaplan, and Stucke [2013].

[2]This convergence fails when $\beta \geq 1$, which means that the time-series is non-stationary ($\beta = 1$ implies that it has a unit root).

differences, which seems undesirable.

We present a new variance-decomposition model of PE performance that better captures the cross-sectional intuition. Our model explicitly captures skill and luck, so an unskilled PE firm with a lucky fund is not immediately considered a skilled firm. We model heterogeneous skills of PE firms, and our model allows some PE firms to consistently outperform. It does not imply that all firms converge in the limit. Separating skill from luck also leads to a natural distinction between two types of persistence, which we term *long-term* and *investable* persistence: Long-term persistence reflects the average outperformance of more skilled PE firms, and it captures how heterogeneous skills affect PE performance. In contrast, investable persistence captures whether investors ("LPs") can identify the skilled PE firms. When performance is random, top-quartile performance may be due to luck, and it does not necessarily imply top-quartile skills. This distinction matters. We find a large amount of long-term persistence, and skilled PE firms outperform by 7% to 8% annually, across all fund types. This performance is noisy, though, and we find only a small amount of investable persistence, particularly for venture capital ("VC") firms. VC performance is mostly due to luck, and an LP needs to observe the performance of an excessive number of past funds (on the order of 25 to 50 past funds) to identify a VC firm as having top-quartile skills with reasonable certainty.

Comparing different subsamples, we find that smaller funds have greater persistence than larger funds. Particularly large VC funds have weak long-term persistence and worse signal-to-noise ratios. Comparing locations of PE firms, we find the least persistence for PE firms located in the US, followed by Europe, and the greatest persistence for PE firm located in the rest of the world ("ROW"), although these PE firms also have more volatile performance. We confirm that persistence has declined in the 2000s relative to the 1990s. This decline is largest for VC firms, and we find that BO and Other funds still show substantial long-term persistence, even post 2000.

Our finding of large long-term persistence but little investable persistence has several implications: First, it explains LPs' increasing focus on obtaining detailed information about PE firms and their past funds (such as the PE firms' internal organization and culture, internal compensation and alignment of incentives, processes and deal sourcing) to help them attribute past performance (e.g., Ewens and Rhodes-Kropf [2013]). Our results show that such detailed information is necessary for LPs to identify top PE firms. Information about past fund performance, by itself, is insufficient. Second, our results may explain why outperformance is not competed away. PE skills are scarce, but when performance is noisy, LPs with the ability to identify skilled PE firms may also be scarce, and those LPs earn rents (Lerner, Schoar, and Wongsunwai [2007] and Berk, Wang, and Weisbach [2013] study heterogeneous LP skills). Last, our findings confirm the economic realities behind the common saying among VCs that "I'd rather be lucky than smart."

Model For our analysis, we develop a variance decomposition model, or *hierarchical linear model*, which generalizes the classical *analysis of variance* ("ANOVA") methods. We estimate our model with a Bayesian procedure, as described in the appendix. The model has several advantages:[3] First, as mentioned, our model distinguishes two different notions of persistence: Long-term persistence, which arises from the difference between skilled and unskilled PE firms, and investable persistence, which reflects the LPs' ability to identify skilled firms. Second, it explicitly models the timing of the funds, and it does not rely on the numbering of these. This is important for simultaneous funds, where it is arbitrary which one is labeled N and $N-1$, and our model distinguishes situations where fund N follows fund $N-1$ by a few months from those where they are years apart. Third, unlike AR(1) regressions, our estimates incorporate the performance of all funds, including those raised by firms with only a single fund (without a fund $N-1$), which are typically weaker firms. Fourth, our model has a non-parametric component, and it does not require

[3]Persistence is also sometimes studied by estimating transition probabilities across fund quartiles (Billingsley [1961] surveys the statistical issues that arise when estimating parameters and testing hypotheses in Markov chains). If funds' performances are *i.i.d.*, then the probability that a top-quartile fund remains top quartile is 25%; and more generally, $P[y_{i,N} \in Q | y_{i,N-1} \in Q] = 25\%$ when Q contains the performance for any quartile. Hence, the empirical finding that $P[y_{i,N} \in Q | y_{i,N-1} \in Q] > 25\%$ implies that performance cannot be *i.i.d.*, which is sometimes interpreted as evidence of persistence, but this interpretation is tenuous. For example, let there be two types of PE firms, with an equal number of each. The first type determines the return of each of its funds by flipping a dime, and it is either +10% or -10%, with equal probability. The second type flips a quarter, and its returns are either +25% or -25%. Hence, the returns for the four quartiles are: +25%, +10%, -10%, and -25%. For each quartile, the transition probability is $P[y_{i,N} \in Q | y_{i,N-1} \in Q] = 50\%$, so returns are not *i.i.d.* (obviously), but there is no persistence in the conventional sense. Conversely, finding transition probabilities of $P[y_{i,N} \in Q | y_{i,N-1} \in Q] = 25\%$, by itself, does not imply an absence of persistence. It is neither a necessary nor sufficient condition. Hence, the economic magnitudes and statistical significance of persistence are difficult to evaluate using transition probabilities across quartiles.

fund returns to be normal distributed, unlike standard ANOVA models. Instead, we use a mixtures-of-normals distribution and a Bayes factor test to determine the appropriate number of mixtures. This generality is especially important for VC funds, which have highly skewed returns. Fourth, our Bayesian approach is computationally efficient, and it provides accurate small sample inferences for the estimated parameters, which is important when the parameters of interest are variances (and a ratio of variances in the signal-to-noise ratio), which have non-standard asymptotic distributions.

Literature Following Kaplan and Schoar [2005], a number of studies have investigated the persistence of PE performance. Phalippou and Gottschalg [2009] consider persistence after correcting for potential biases in reported interim NAVs. Chung [2010] and Phalippou [2010] find weaker effects when regressing $y_{i,N}$ on $y_{i,N-2}$, and they argue that persistence is short lived. Hochberg, Ljungqvist, Vissing-Jorgensen [2014] model performance persistence that arises from asymmetric information between the LP and GP. Recently, Harris, Jenkinson, Kaplan, and Stucke [2013] find that persistence has declined post 2000 for BO firms, and this finding is confirmed by Braun, Jenkinson, and Stoff [2013], using deal-level data. Separating skill from luck is a general question in economics and finance, and our analysis may be useful for other applications, such as the persistence of the performance of serial entrepreneurs (e.g., Gompers, Kovner, Lerner, and Scharfstein [2010] and Bengtson [2013]).

Outline The paper proceeds as follows. In Section I we present the data. Section II presents our empirical model. Section III presents our results and discusses the evidence for "long-term" persistence in private equity performance. Section IV evaluates the "investable persistence." Section V analyzes various subsamples of the data, and Section VI concludes. We provide a detailed description of the Bayesian estimation procedure in the Appendix.

I Data

This paper uses an extensive dataset with PE firms, the funds they manage, and the performance and other information for these funds. The data are obtained from Preqin, a commercial data provider that started collecting performance data using freedom of information act ("FOIA") requests to public investors and later extended the scope of its data collection to other public filings and voluntary reporting by some GPs and LPs. For each fund, Preqin only reports aggregate fund performance, such as the IRR and Total Value to Paid-in Capital multiple ("TVPI"). We do not have individual cash flows between the LPs and GPs. A limitation is that the data do not contain the public market equivalent ("PME") measure of fund performance, which has advantages when evaluating risk-adjusted performance (see Sorensen and Jagannathan [2013] and Korteweg and Nagel [2013]).

Harris, Jenkinson, and Kaplan [2013] compare several datasets with PE fund performance. Most of these data are from commercial data providers (Preqin, Burgiss, and Cambridge Associates) and one is from a large anonymous LP (studied by Robinson and Sensoy [2011]). For buyout ("BO") funds, they find that Preqin contains the largest total number of funds in the 1990s and 2000s (but not in the 1980s). For venture capital ("VC") funds, Preqin has slightly weaker coverage in the 1980s and 1990s, but it is the most comprehensive dataset in the 2000s. Importantly, of all the datasets, the Preqin data contain performance information for the greatest number of both BO and VC funds. Moreover, they find no evidence that Preqin's performance data are biased relative to the performance data from other data sources. Hence, when analyzing the performance and persistence of PE funds, the Preqin data are among the best data sets currently available.

The Preqin data contain information about each fund's type. The two main fund types are buyout ("BO") and venture capital ("VC") funds, but Preqin also classifies funds as: real-estate, fund-of-funds, infrastructure, turn-around, special situations, co-investment, and venture debt funds, which we collectively refer to as "Other" funds. The majority of "Other" funds are real-estate and infrastructure funds, and while these two fund types are quite different, we find that they have (surprisingly) similar performance and persistence, and we combine all of these fund types for most of our analysis. We define a fund's geographical location by the location of its GP. This location may differ from the locations of the portfolio companies, but we obtain very similar results when we instead define location in terms of the fund's geographical investment focus.

Sample We restrict our sample to funds with available performance information. Our model explicitly captures the timing of the individual funds. It does not rely on the numbering of funds, and our estimates are valid even when performance data are missing for some (randomly chosen) funds. We avoid concerns raised in recent studies about funds' self-reported intermediate IRRs, TVPIs and NAVs ("net-asset values"), by restricting our sample to fully liquidated funds. Finally, we restrict our sample to funds with at least USD 5M of committed capital (in 1990 dollars) to exclude smaller idiosyncratic funds. Our final sample contains 1,924 funds, raised between 1969 and 2001, and managed by 891 firms. There are 842 venture capital ("VC") funds, 562 buyout ("BO") funds, and the remaining 518 funds are classified as Other funds. Table I shows summary statistics for our final sample. Panel B shows sub-classifications of VC and Other funds.

** TABLE I: SUMMARY STATISTICS **

Internal Rate of Return Preqin reports each fund's internal rate of return ("IRR"). The IRR is the annualized return to the limited partners ("LPs") in the fund, net of performance fees ("carried interest" or "carry") and management fees. While the IRR has well-known limitations, it is the most widely available fund performance measure and commonly used in studies of fund performance.

** TABLE II: FUND IRRs BY VINTAGE YEAR **

Table II reports the average IRR for each year. These IRRs are plotted in Figure 2 for VC, BO, and Other funds. For VC funds, we see strong performance during the dot-com bubble in the late 1990s, with average (annualized) IRRs as high as 45.2%, followed by the sharp drop after the bursting of the dot-com bubble. Each fund has a ten-year life, and the indicated year is the fund's year of inception ("vintage year"), so funds with vintage years well before 2000 were exposed to the bubble and show lower performance. BO performance has been more stable, and it has recently shown a strong recovery relative to VC and Other funds. The performance of Other funds has been even more stable, showing an earlier but more modest decline in the late 1990s, followed by a corresponding recovery.

** FIGURE 1: IRRs BY VINTAGE YEAR **

Our analysis uses total log-returns (or "continuously compounded" returns) rather than annualized IRRs, which is reported by Preqin. The total (log-)return for fund u is denoted y_{iu}, and it is calculated by compounding the fund's IRR over its ten year life, as follows:

$$y_{iu} = 10 \cdot \ln(1 + IRR_{iu}) \,. \qquad (2)$$

This calculation fails for two funds that have IRRs of -100% (one is a 2001 VC fund and the other is a 1998 BO fund). Our analysis excludes these two funds, but our results are robust to including them with IRRs set equal to the first (lowest) percentile of the IRR distribution.

II Variance Decomposition

For our empirical analysis, we use a *hierarchical linear model*, which generalizes the classical *analysis of variance* ("ANOVA") decomposition. Hierarchical models, using Bayesian estimators that exploit advances in numerical computing (Markov-Chain Monte Carlo, Gibbs sampling, and posterior augmentation), have recently been extensively developed and applied. These models were initially used for educational measurement, because they capture the hierarchical structure that arises when, for example, one observes individual students, who are grouped into classrooms, in different schools, in different districts, etc. (For introductions to hierarchical models and more applications see Raudenbush and Iryk [2008] and de Leeuw and Meijer [2008].) This hierarchical structure also arises for PE when individual PE funds are managed by different PE firms and span different time periods (with data for individual deals, as in Braun, Jenkinson, and Stoff [2013] or with LPs' holdings of PE funds, as in Sensoy, Wang, and Weisbach [2013], our model can be extended to include data at these additional levels as well).

Modeling the hierarchical structure avoids the *unit of analysis* problem (Burstein et al. [1980]). When studying the persistence of PE performance, we are interested in differences between PE firms, so the unit of analysis is a PE firm, but the unit of observation is the underlying funds, which are *repeated measures* of the PE firm's quality. Increasing the number of funds per firm improves the estimate of each firm's quality but not the number of firms that are compared. With few firms but many funds per firm, observing even more funds per firm becomes uninformative, because the main sampling error arises from the sampling of the firms' qualities, not the sampling of the funds observed for each firm. In contrast, increasing the number of observed firms always improves the estimates. It is difficult for classical regression models, using PE firm-fixed effects ("FEs"), to address this problem, because these models only consider the sampling of the funds for a given set of PE firms (i.e., a given set of PE-firm FEs), not the sampling of the PE firms themselves (i.e., the sampling of the observed FEs from a larger population of potential FEs).

Economic Intuition To illustrate the intuition behind our variance decomposition, consider 60 PE firms. Each firm makes two investments (or manages two funds), each of which either succeeds or fails. For the resulting 120 investments, say, we observe that one half fails and the other half succeeds, so the unconditional success probability is 50%. If the individual investments were statistically independent, each investor would have 25% probability of zero successful investments, 50% probability of a single success, and 25% chance of two successes. We would then see 15 of the 60 PE firms with no successes, 30 with a single one, and the remaining 15 firms with two successful investments. Imagine instead that the observed successes are evenly distributed among the 60 PE firms, so 20 have zero, 20 have one, and 20 PE firms have two successes. In other words, the performance variation *between* PE firms exceeds the amount of variation that is implied by the investments *within* PE firms, if the investments were independent. In this case, the investments cannot be independent, obviously, so some PE firms must have higher (and lower) success probabilities. In other words, some PE firms persistently show better (and worse) performance. For example, the even distribution of success among PE firms is consistent with each PE firm's success probability being drawn from the uniform distribution on $[0,1]$. If p_i denotes firm i's success probability, then the expected probability of two successes is $E[p_i^2] = 33\%$ when $p_i \sim U[0,1]$. Based on this intuition, we define and measure persistence by comparing the performance variability *within* funds to the performance variability *between* PE firms. When there is excess variation between firms, as in this example, it implies persistence.

This intuition also leads to a natural distinction between PE firms with high skill and high performance. With $p_i \sim U[0,1]$, using Bayes rule, conditional on observing two successes, the posterior density of p_i is $f(p_i|SS) = 3p_i^2$. Hence, the probability that a firm with top-tercile performance (two successes) has top-tercile skill is $\Pr(p_i \in [0.66;1]|SS) = 70\%$. And the expected success probability for a subsequent fund by a firm with top-tercile performance is only $E[p_i|SS] = 75\%$, whereas the success probability for a firm with actual top-tercile skill is $E[p_i|p_i > 66\%] = 83\%$.

Note that performance is a noisy indicator of skill even when the skill distribution is perfectly known (i.e., $p_i \sim U[0,1]$ is known). When the skill distribution is estimated, additional uncertainty arises due to estimation error. Our model, which we discuss next, incorporates this parameter uncertainty as well.

Formal Model Let PE firms be indexed by i. Each PE firm manages a sequence of underlying PE funds, indexed by u. Each observation contains the fund's performance and other characteristics of the firm and fund. We specify the ten-year total log-return of fund u as:

$$y_{iu} = X_{iu}'\beta + \sum_{\tau=t_{iu}}^{t_{iu}+9} (\alpha_i + \eta_{i\tau}) + \varepsilon_{iu}\,. \tag{3}$$

Here, X_{iu} contains time fixed effects for the timing of the fund (formally, the model is then a *mixed-effects model*). The sum runs over the funds' ten-year life, with year t_{iu} denoting the fund's first year of operations ("vintage year"). The three terms α_i, $\eta_{i\tau}$, and ε_u are three random effects that define the variance-covariance structure across the funds' performances. Our model cannot determine when a given fund's return is earned during its life, because we only observe each fund's ultimate performance. The model can determine, however, how much of the variation in this performance that is due to each of the three random effects.

Statistical Properties The random effects in equation (3) decompose the variation in fund performance into three parts: First, α_i is the PE-firm effect, reflecting long-term persistence. For each PE firm, it is distributed $\mathcal{N}(0,\sigma_\alpha^2)$, and it is constant for all funds managed by the same firm. We interpret a PE firm with a high α_i as having greater skills (corresponding to a higher success probability, p_i, in the example). The model is parameterized with α_i inside the sum in equation (3), so each fund "earns" α_i ten times, and α_i is the annualized return to the PE firm's skill. Second, $\eta_{i\tau}$ is the PE firm-time effect. For each firm and year, $\eta_{i\tau}$ is distributed *i.i.d.* $\mathcal{N}\left(0, \sigma_\eta^2\right)$. Two partially overlapping funds that are managed by the same firm will share an $\eta_{i\tau}$ term for each year of overlap, which introduces correlation between partially overlapping funds that are managed by the same firm. Third, ε_u is an error term, capturing the residual idiosyncratic variation in each fund's performance. Because fund performance is highly skewed, we allow ε_u to be distributed as a mixture of normals, which is considerably more flexible than the normal distribution.[4]

The sum in equation (3) contains the same α_i term ten times, and ten *i.i.d.* $\eta_{i\tau}$ terms, so the total variance of y_u is:

$$\sigma_y^2 = 100\sigma_\alpha^2 + 10\sigma_\eta^2 + \sigma_\varepsilon^2 . \tag{4}$$

Economic Motivation The three random effects are motivated as follows: First, some PE firms may have particular investment or management skills that improve the performance of all their funds. Such long-term persistence is captured by the α_i term, the variation in α_i across PE firms captures differences in skills across PE firms. When there is little variation in α_i, corresponding to a small σ_α^2, then PE firms are similar, and there are few persistent differences in their performance. When σ_α^2 is large, more of the performance difference is due to heterogeneous skills of the PE firms.

Second, PE firms typically manage several contemporaneous funds. During the overlap period, these contemporaneous funds have common market exposures, which introduces correlations in their performances. To illustrate, a PE firm that manages two funds with vintage years 1999 and 2001 may be focusing on investments in emerging markets. Hence, during 2001–09, the two funds will be exposed to similar emerging-market shocks. In this case, a regression of $y_{i,N}$ on $y_{i,N-1}$ yields a positive coefficient, but this coefficient is not evidence of persistence, as usually defined. It does not imply that a firm's past performance predicts its future performance. Instead, it arises from the spurious correlation due to the

[4] Using Bayes factors to test model specifications, we find that VC performance requires a mixture of three normals whereas the performance of Buyout and Other funds are captured by mixtures of one or two normal distributions.

funds' unobserved shared exposures. In our model, these shared components are captured by $\eta_{i\tau}$. All overlapping funds that are managed by the same PE firm share an $\eta_{i\tau}$ term for each year of overlap. These shared terms capture the increasing correlation between funds with greater overlaps. When the estimated σ^2_η is large, this overlap effect is large. Formally, the covariance between two funds that are managed by the same PE firm, with N years of overlap, is:

$$\mathrm{COV}(y_{iu}, y_{iv}) = 100\sigma^2_\alpha + N\sigma^2_\eta. \tag{5}$$

This covariance relationship is plotted in Figure 2, and this figure illustrates the identification of the model. The main parameters of interest are the variances of the three random effects, σ^2_α, σ^2_η, and σ^2_ε. In Figure 2, the intercept is σ^2_α and the slope is σ^2_η, so these two variances are identified by comparing the covariances of funds with increasing amounts of overlap. Given σ^2_α and σ^2_η, and observing total variance, σ^2_y, the residual variance in equation (4) identifies σ^2_ε.

** FIGURE 2: OVERLAP AND COVARIANCE **

III Results

A IRR Regressions

We first confirm the original findings by Kaplan and Schoar [2005] using our data. Table III reports coefficients from OLS regressions of $IRR_{i,N}$ on $IRR_{i,N-1}$, and the reported coefficients show that the previous fund's performance strongly predicts the performance of the subsequent fund. In Specification I, the positive and significant coefficient of 0.125 suggests that a VC fund with a 1% higher IRR predicts a 0.125% higher IRR for the subsequent fund. Specification II suggests that this effect is even stronger when controlling for the performance of fund $N-2$, although the coefficient on this second fund's performance is negative. For BO funds we find slightly stronger positive and significant effects. For Other funds, the coefficient is positive and significant in Specification I, but it becomes smaller and insignificant when including fund $N-2$ in Specification II, although this weaker statistical result may be due to the smaller sample size.

In these specifications, even fund $N-2$ may still overlap with fund N, and the positive coefficients may reflect this overlap rather than actual persistence. In Panel B of Table II, we reduce the sample to funds that are entirely non-overlapping, which further reduces the sample size and leaves no remaining signs of persistence, although this weaker result may just reflect the lower statistical power due.

Finally, note that none of the specifications in Table III suggest that performance is systematically related to fund size, and there is some weak evidence that a higher sequence number is associated with better performance.

** TABLE III: IRR REGRESSIONS **

A natural interpretation of the results from the AR(1) regression is that BO funds have the most persistence (largest coefficients and R^2, and significant with fund $N-2$), followed by VC funds (smaller coefficients and R^2 than BO funds, but significant with fund $N-2$), and that Other funds show the least, if any, performance persistence (smallest coefficients and R^2, and insignificant with fund $N-2$). This analysis, however, does not distinguish skill from luck, and it does not distinguish long-term from investable persistence.

B Long-Term Persistence

Table IV reports the estimated parameters in our model. Panel A shows the magnitudes of the three random effects as measured by their standard deviations (σ_α, σ_η and σ_ε).[5] The decomposition of the variances ($100 \times \sigma_\alpha^2$, $10 \times \sigma_\eta^2$, σ_ε^2 and σ_y^2) is easier to interpret, and is reported in Panel B.

** TABLE IV: PARAMETER ESTIMATES **

[5]We use a Bayesian estimator, but we report results using standard frequentist terminology: The "point estimate" is the mean of the posterior distribution, and the "standard error" is the the standard deviation of the posterior distribution. A parameter is "statistically significant," at a given level, when zero is not contained in the corresponding symmetric credible interval, as usually defined in Bayesian statistics. Our Bayesian estimator produces exact small-sample inference, even for non-linear transformations of the estimated parameters, and all reported inference is calculated this way. We do not rely on any asymptotic approximations.

Buyout For BO funds, Specification I of Table IV shows a total unconditional variance (σ_y^2) of 2.428. This variance can be decomposed into three effects, with 0.361 due to long-term persistence ($100 \times \sigma_\alpha^2$), 0.216 due to the overlap effect ($10 \times \sigma_\eta^2$), and the remaining 1.852 due to idiosyncratic variance (σ_ε^2). The long-term persistence effect, as measured by σ_α, is statistically significant,[6] consistent with the earlier findings using the AR(1) regression.

To evaluate the economic magnitude of the long-term persistence, note that the annual contribution of a PE firm's skill is α_i, which is distributed $\mathcal{N}\left(0, \sigma_\alpha^2\right)$, and let $q_\alpha(\cdot)$ denote the percentiles of the α_i distribution. Using the point estimate of σ_α of 0.060, we have $q_\alpha(50\%) = 0$ and $q_\alpha(75\%) = 4.05\%$, so the median firm has an alpha of zero, by definition, and the marginal (worst) top-quartile BO firm has an α_i of 4.05%, annually. The median top-quartile BO firm has an α_i of $q_\alpha(87.5\%) = 6.90\%$. Hence, the spread between the marginal top- and bottom-quartile firms, due to skill, is $q_\alpha(75\%) - q_\alpha(25\%) = 8.09\%$, annually. This calculation, however, assumes that the skill distribution is perfectly estimated. In Table IV, Specification I for Buyout funds shows a standard error of 0.008 for the σ_α estimate. To account for this estimation error, note that our estimation procedure simulates the full posterior distribution of σ_α, and we can calculate the corresponding posterior distribution of $q_\alpha(75\%) - q_\alpha(25\%)$. The mean of this posterior distribution, which accounts for the estimation error in the skill distribution, is 7.93%, as reported in Table IV. This estimate of 7.93% is close to 8.09%, which is calculated using just the point estimate of σ_α without adjusting for estimation error.[7] Hence, adjusting for estimation error in the skill distribution seems to have a minor effect on the estimate of the long-term persistence. Nevertheless, because the adjustment is simple, all reported alpha spreads in Table IV adjusts for this estimation error in the skill distribution. In addition to our estimate of the interquartile range, $q_\alpha(75\%) - q_\alpha(25\%)$, of 7.93%, annually, we also report our estimates of the performance difference between the median top- and bottom-quartile firms, denoted $q_\alpha(87.5\%) - q_\alpha(12.5\%)$. Our estimate of this difference is 13.63%, annually.

[6] Testing statistical significance of variance parameters is complicated by the one-sided alternative hypothesis. We use a Bayes factor test to test $H_0 : \sigma_\alpha^2 = 0$ against $H_A : \sigma_\alpha^2 > 0$, as reported in Table VII and discussed in the Appendix.

[7] Accounting for estimation error in the skill distribution reduces the estimated spread in alphas, because estimator error concentrates mass at the center of the skill distribution.

Note that these alpha spreads cannot be calculated as the empirical difference between the IRRs of top- and bottom-quartile funds, because top-quartile performance does not imply top-quartile skills, so this empirical difference confounds skill and luck. If σ_ε^2 is large, but σ_α^2 is zero, there is no long-term persistence, and $q_\alpha(75\%) - q_\alpha(25\%)$ is zero. But a large σ_ε^2 still implies a large difference in fund performance, albeit due to noise, so the empirical difference would still be large, and it would overstate the performance that is due to heterogeneous skills. The empirical difference may also understate long-term persistence. In periods where a disproportionate number of high-quality (or low-quality) firms are active, the empirical difference may be too small, because it is calculated from funds in a narrow range of the α_i distribution. For this reason, it is also important that our model accommodates firms that only raise a single fund. These firms are likely from the lower tail of the α_i distribution, and excluding them would introduce a downward bias in σ_α^2 and underestimate the long-term persistence.

Venture Capital For VC funds, the variance that is due to long-term persistence ($100 \times \sigma_\alpha^2$) is 0.243, which is similar to BO funds, and therefore the alpha spreads are also similar. Specifically, for VC firms, $q_\alpha(87.5\%) - q_\alpha(12.5\%) = 11.17\%$ and $q_\alpha(75\%) - q_\alpha(25\%) = 6.50\%$, annually. The variance due to the overlap effect ($10 \times \sigma_\eta^2$) is 0.675, which is somewhat larger than for BO funds, but this difference disappears with vintage-year FEs. Importantly, VC funds have much greater idiosyncratic risk than BO and Other funds. Hence, even though the alpha spreads are are similar, the performance of VC firms is much more noisy. This noise may also explain the weaker persistence results for VC funds using the AR(1) regression, because more noisy outcomes results in weaker statistical power in this model.

Other For Other funds, the overlap and long-term persistence effects are similar to those for BO and VC funds. Comparing the first specifications in Table IV, the σ_α estimates are close for Other and VC funds, so their alpha spreads are also almost identical. The idiosyncratic volatility, however, is lower for Other funds.

Overall, the long-term persistence, is greatest for BO funds and slightly lower for VC and Other funds, but the differences are small, in economic magnitude. Moreover, these alpha spreads are calculated from the underlying distribution of the firms' skills. For an LP to earn the this spread, this LP must be able to perfectly discriminate between skilled and unskilled firms. Hence, these spreads represent upper bounds on the returns that LPs can earn by investing in skilled relative to unskilled PE firms.

C Overlap Effect

The performances of overlapping funds is correlated, mechanically, even for funds that are managed by different PE firms. In Table IV, Specifications I and II show overlap effects of 0.675 and 0.386 for VC funds. Without vintage-year FEs, the $\eta_{i\tau}$ terms capture all correlations between contemporaneous funds, including the correlation that is due to their shared market exposure during the overlap period. To control for these shared exposures, Specification II includes vintage-year FEs. With these FEs, the overlap effect captures the correlations between overlapping funds that are specific to the funds that are managed by an individual PE firm. Comparing the magnitudes, BO funds have the largest and Other funds have the smallest overlap effect. For both BO and VC funds, the variation due to the overlap exceeds the variation due to long-term persistence (in Specification II, with vintage-year FEs).

The overlap effect is important, because the AR(1) analysis that regresses the performance of fund N on $N-1$ will find a positive coefficient due to this effect, but this coefficient does not reflect persistence in the conventional sense. The $\eta_{i\tau}$ terms that generate this effect are purely transitory, they are *i.i.d* over time, and the correlation between subsequent funds does not indicate that past performance predicts future performance.[8] Table I shows an average overlap of subsequent funds of 5.8–6.8 years. Using equation (5), the estimates in Table IV imply a total covariance between funds with average overlaps of 0.37–0.64. But 25.8%–61.8% of this covariance is due to the overlap, suggesting that the AR(1) coefficients may be upward biased by 34%–168%. The overlap effect for Other funds is smaller, implying a smaller upward bias in the AR(1) regression, which may partially explain why the AR(1) regression in Table III shows relatively weaker persistence for these Other funds.

[8]Although operating over much longer time scales, the autocorrelation due to the overlap of contemporaneous funds is closely related to the autocorrelation that arises from nonsynchronous trading, which was introduced by Fisher [1966] and has since been extensively studied.

IV Learning and Investable Persistence

The previous section considered long-term persistence, defined as the outperformance by funds that are managed by skilled PE firms. In practice, however, it is difficult for LPs to identify skilled firms. For example, Phalippou [2010] finds that the previous fund's interim performance, which is typically the only performance of this fund that is known when the subsequent fund is raised, is not statistically significant for predicting the performance of the subsequent fund. In other words, the interim performance of the previous fund, by itself, is insufficient for LPs to identify skilled PE firms. This finding raises a more general question: For an LP to evaluate an investment in a new fund, how much information does the LP need in order to determine the PE firm's skill? When LPs need little information, such as just the interim performance of the previous fund, it is easy to identify skilled firms, and PE performance has large investable persistence. Conversely, as we find below, when LPs require much information, skilled firms are difficult to identify, and there is little investable persistence.

We quantify the investable persistence in two ways. First, we estimate the signal-to-noise ratio. This ratio is simple to calculate, it allows for a direct comparison of different types of firms, and it has a simple economic intuition based on the updating of the LPs' beliefs about a PE firm's skills. The disadvantage is that this ratio does not reflect the full statistical model. Consequently, we also use the full model to estimate how many past funds an LP must observe to assess a PE firm's skill with reasonable certainty. Overall, we find that the signal-to-noise ratio is low, and it is difficult for LPs to identify skilled PE firms based on their past performance alone. An LP needs to observe an excessive number of past funds to evaluate an PE firm's skill with reasonable confidence. In practice, this means that LPs need additional information, such as detailed information about individual deals, individual partners associated with these deals (see Ewens and Rhodes-Kropf [2013]), or other additional information. Past performance, by itself, is insufficient for evaluating PE firms.

A Signal-to-Noise

Our model has two types of shocks: Transitory shocks are drawn independently each period, as given by the $\eta_{i\tau}$ and ε_u terms. Persistent shocks, which reflect the heterogeneous skills of PE firms, are given by α_i. It is common to define the signal-to-noise ratio, s_α, as the ratio of the variance due to persistent shocks relative to the total variance:[9]

$$s_\alpha = \frac{100\sigma_\alpha^2}{\sigma_y^2}. \tag{6}$$

This ratio has a simple economic interpretation. In a Gaussian learning model, an LP would update its beliefs about α_i as follows: Let the LP's beliefs about α_i after observing N funds be distributed $\mathcal{N}\left(\alpha_{i,N}, \sigma_{i,N}^2\right)$. After observing the performance of one additional fund, the LP's updated beliefs become $\mathcal{N}\left(\alpha_{i,N+1}, \sigma_{i,N+1}^2\right)$ with:

$$\alpha_{i,N+1} = s_\alpha \times \frac{y_{i,N+1} - X'_{i,N+1}\beta}{10} + (1 - s_\alpha) \times \alpha_{i,N}, \tag{7}$$

and

$$\sigma_{i,N+1}^2 = (1 - s_\alpha) \times \sigma_{i,N}^2. \tag{8}$$

Hence, in this Gaussian learning model, the signal-to-noise ratio shows how much weight an LP places on new information and how fast the precision of the updated beliefs improves. When the ratio is low, new performance is uninformative about the firm's skills, and it is difficult to learn α_i. When s_α is larger, LPs learn faster, as measured by the smaller σ_{N+1}^2.

Figure 3 plots the posterior distributions of s_α for VC, BO, and Other firms, with and without vintage-year FEs. The location of the posterior distribution reflects the estimated value of s_α (the point estimates of s_α are reported in Table VI). The estimated ratio is lowest for VC fund, and then increases for BO and Other funds. For VC funds, the large transitory variance means that relatively less of the variation in VC fund performance is due to persistence, and VC performance is relatively uninformative about the skills of VC firms. Other funds have the least transitory variance, and the best signal-to-noise ratio,

[9] For example, Cochrane [1988] uses a similar variance ratio to evaluate the persistence of GDP shocks.

making it is easier for LPs to identify skilled Other firms.

** FIGURE 3: ESTIMATES OF SIGNAL-TO-NOISE RATIO **

B Identifying Skilled PE Firms

Figure 4 plots $\mathrm{P}\left[\alpha_i \geq q_\alpha(75\%) \,|\, \frac{1}{N}\sum_{n=1}^{N} y_{i,n} \geq Q_N\right]$, which is the probability that a PE firm has top-quartile skills (i.e., $\alpha_i \geq q_\alpha(75\%)$) when its average performance is in the top quartile of observed performance among firms with N past funds (i.e., $\frac{1}{N}\sum_{n=1}^{N} y_{i,n} \geq Q_N$, where Q_N denotes the average performance of the marginal top-quartile firm with N past funds). To interpret Figure 4, consider first the limit case where there is a vanishing amount of long-term persistence and σ_α^2 converges to zero. In this case, top-quartile performance is entirely due to luck, and the probability that a firm with top-quartile performance also has top-quartile skills is just 25%. This uninformative case gives the lower bound on this probability, and as more information becomes available, this probability increases. In the other limit, when σ_α^2 becomes large (relative to σ_y^2), persistence dominates, firms with top-quartile skills are perfectly identified by their top-quartile performance, and the probability approaches 100%. In practice, for any reasonable number of funds, the probability remains well below this upper limit. To illustrate, for VC firms with five past funds, top-quartile skills only conveys a 37% probability of showing top-quartile performance, which seems only slightly better than their 25% probability in the uninformative case, without any persistence. Of the BO and Other firms with five past funds and top-quartile performance, 47% and 51% have top-quartile skills. These results are consistent with the estimates of the signal-to-noise ratio, which also found that Other funds have the most informative performance, followed by BO and VC funds.

Figure 4 shows probabilities for up to 50 past funds. Since no current PE firm has managed 50 fully liquidated funds, this represents an upper bound on the ability of LPs to discriminate between PE firms based on their past performance alone. Even at this upper bound, just 53% of the VC firms with top-quartile performance actually have top-quartile skills.

** FIGURE 4: LEARNING SPEED **

C Investable Persistence

The previous results show that skilled Other firms are most easily identified, followed by BO and VC firms, but they do not account for the value of identifying skilled firms. In Table IV we saw that BO firms have the highest long-term persistence, measured by σ_α^2, and hence the greatest performance difference between skilled and unskilled BO firms, measured by their alpha spreads. Hence, although more difficult, it is also more valuable to identify a skilled BO firm. Figure 5 shows the combined effect of the difficulty of identifying skilled firms and the value of identifying them. This figure plots the expected alpha when of a firm with top-quartile performance, where top-quartile performance is calculated among PE firms with a given number of past funds, N. Formally, Figure 5 plots $E\left[\alpha_i \,|\, \frac{1}{N}\sum_{n=1}^{N} y_{i,n} \geq Q_N\right]$. To interpret this figure, note that while skilled Other firms can be more accurately identified, the value of investing in them is limited by their alpha spread. When only a few past funds are observed, this ability to identify the skilled Other firms dominates. For BO firms, however, the alpha spread is greater, and after observing 4–5 past funds, skilled BO firms are sufficiently well identified that the benefit of their greater spread outweighs the difficulty of identifying them. In this case, BO funds have the greatest investable persistence.

** FIGURE 5: INVESTABLE PERSISTENCE **

VC firms have poor investable persistence overall. Their signal-to-noise ratio is low, and it is difficult to identify skilled VC firms. Moreover, VCs' alpha spread is modest, it is similar to the spread of Other firms, and well below the spread of BO firms. The resulting weak investable persistence of VC firms shows in Figures 4 and 5, where VC firms place substantially below Other and BO firms.

V Subsamples

Table VI presents estimates of our model for different subsamples of our data. We calculate these estimates by dividing our sample into separate subsamples, and estimating the model on each individual subsample. For example, when we show the persistence of small and large funds in Panel A of Table VI, the model is estimated for each subsample separately, so PE firms that manage both small and large funds are represented, independently, in both of estimates.

** TABLE VI: SUB-SAMPLES **

Fund Size Table VI first compares the long-term persistence of small and large funds, and it shows that smaller funds have long-term persistence and greater alpha spreads than larger funds, consistently across VC, BO, and Other funds. Hence, for smaller funds, there is a greater performance difference between those managed by skilled and unskilled firms. Note that this performance difference is not due to larger idiosyncratic volatility of smaller funds. Smaller funds do have larger volatilities (except for Other funds), but this volatility is captured separately by the σ_ε^2 terms. Panel B of Table VI shows that it is also easier to identify skilled firms from the performance of smaller funds. Despite their more volatile performance, the estimated σ_α^2 parameters are sufficiently large that the overall signal-to-noise ratios are still better for smaller funds. Hence, smaller funds have substantially greater persistence, both in terms of long-term and investable persistence.

GP Location Table VI shows that PE firms located in the rest of the world ("ROW") have more long-term persistence, followed by firms located in Europe, and US-based firms have the least long-term persistence. Total volatility follows a different pattern, with ROW being most volatile, followed by the US and then European-based funds. For VC and BO funds this results in a signal-to-noise ratio that is substantially better for European-based funds. For Other funds, it is funds located in ROW that have the most informative performance. The performance of US-based funds is relatively uninformative, and skilled US-based firms are more difficult to identify, which is consistent with the US PE industry being more mature.

Investment Style VC and Other funds can be further classified by their investment style, as given by Preqin. Table VI shows that VC firms that focus on early-stage investments have the worst long-term performance and the least informative performance. Generalist

VC firms show slightly more long-term persistence, but Late-stage VC firms show slightly more informative performance.

For Other funds, we can distinguish Real-estate from Fund-of-funds. These two types of funds are very different, but they have surprisingly similar persistence characteristics. Fund-of-funds have slightly greater long-term persistence than Real-estate funds, although their long-term persistence is still well below the levels of VC and BO funds. But Real-estate funds have more informative performance than Fund-of-funds, and their performance is more informative than that of both VC and BO funds.

Time Period Finally, we confirm the finding by Braun, Jenkinson, and Stoff [2013] and Harris, Jenkinson, Kaplan, and Stucke [2013] that persistence has declined. Table VI shows our persistence estimates for the early and late half of our sample period. Panel A shows that long-term persistence has declined substantially across all fund types. Panel B shows that fund performance has also become less informative about the skills of the PE firms, although this decline is particularly pronounced for VC firms. Overall, for BO and Other firms, there seems to be substantial remaining spreads in the performance of skilled and unskilled firms, and these skilled firms have only become slightly more difficult to identify in the later part of our sample period. This finding is also consistent with the increasing focus by LPs on collecting additional information about the PE firms and their underlying funds.

VI Conclusion

We decompose private equity ("PE") performance into skill and luck. When performance is noisy, top-quartile performance does not necessarily imply top-quartile skills. This distinction leads to two new notions of performance persistence: First, long-term persistence reflects the performance differences between funds managed by skilled and unskilled PE firms. Across all types of PE firms, we find a large amount of long-term persistence, and skilled PE firms outperform by 7% to 8%, annually. Second, investable persistence reflects the ability of LPs to identify skilled PE firms from their past performance. We find that past performance is noisy, with a poor signal-to-noise ratio, making it difficult to identify skilled PE firm, particularly for VC firms. An LP needs to observe an excessive number of past funds to identify those with top-quartile skills with reasonable certainty. For example, even after observing 50 past funds, PE firms with top-quartile performance only have 53%-61% probability of also having top-quartile skills. In practice, to identify skilled PE firms, LPs need information that go beyond the performance of their past funds.

Subsamples Comparing subsamples, we find that smaller funds have greater persistence than larger funds. In particular, large VC funds have poor long-term and investable persistence. Across geographical locations of PE firms, we find the least persistence for PE firms located in the US, followed by Europe, and the greatest persistence for PE firm located in the rest of the world ("ROW"), although these PE firms also have more volatile performance. Finally, we confirm the finding by Harris, Jenkinson, Kaplan, and Stucke [2013] and Braun, Jenkinson, and Stoff [2013] that persistence has declined in the 2000s relative to the 1990s. This decline is largest for VC firms, though, and we find that Buyout and Other funds still show substantial long-term persistence, even post 2000.

Implications Our results have three implications for understanding PE performance: First, the low investable persistence explains LPs' increasing interest in collecting more detailed information about PE performance. For example, Ewens and Rhodes-Kropf [2013] study performance using deal- and partner-level information. We show that such detailed information is necessary for LPs to identify top PE firms. Second, our results provide an alternative explanation of why persistent outperformance is not competed away. When identifying skilled PE firms is difficult, LPs with this ability may also be scarce, and LPs with this ability should earn rents. Third, the large idiosyncratic risk and lower skill components of VC performance is consistent with the anecdotal saying among VCs that "I'd rather be lucky than smart."

Appendix: Estimation Procedure

We implement the model as a Bayesian multi-level hierarchical model, redefining the error terms to absorb the firm-specific random effects using hierarchical centering, as recommended by Gelfand, Sahu, and Carlin [1995]. The performance of fund u of firm i is:

$$y_{iu} = X_{iu}\beta + \sum_{\tau=t_{iu}}^{t_{iu}+9} \eta_{i\tau} + \varepsilon_{iu}, \tag{A.1}$$

The conditional distributions of the random effects are given as:

$$\eta_{i\tau}|\alpha_i \sim \mathcal{N}\left(\alpha_i, \sigma_\eta^2\right), \tag{A.2}$$

$$\alpha_i \sim \mathcal{N}\left(0, \sigma_\alpha^2\right). \tag{A.3}$$

The fund-specific error term distribution is IID

$$\varepsilon_{iu} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right). \tag{A.4}$$

We are interested in estimating the parameter vector $\theta \equiv \left(\beta, \sigma_\alpha^2, \sigma_\eta^2, \sigma_\varepsilon^2\right)$, given a dataset of fund returns, $\{y_{iu}\}$, the dates of inception and termination of each fund, and the set of observed fund-level covariates, X_{iu}. We augment the parameter vector with the latent α's and the $\eta's$, and use a Bayesian estimation algorithm that produces a set of draws from the posterior distribution, $f(\theta, \{\alpha_i\}, \{\eta_{it}\} | data)$, using a Gibbs sampler (Gelfand and Smith [1990]. See also Korteweg [2013] for a detailed description). By the Hammersley-Clifford theorem, we can divide the posterior into five *complete conditionals* that are easy to sample from:

1. Latent firm-year random effects: $f(\{\eta_{it}\} | \{\alpha_i\}, \theta, data)$

2. Variance of fund-specific error term and β-coefficients: $f\left(\sigma_\varepsilon^2, \beta | \{\alpha_i\}, \{\eta_{it}\}, \sigma_\alpha^2, \sigma_\eta^2, data\right)$

3. Latent firm random effects: $f(\{\alpha_i\} | \{\eta_{it}\}, \theta, data)$

4. Variance of firm-year random effects: $f\left(\sigma_\eta^2 | \{\alpha_i\}, \{\eta_{it}\}, \beta, \sigma_\alpha^2, \sigma_\varepsilon^2, data\right)$

5. Variance of firm random effects: $f\left(\sigma_\alpha^2 | \{\alpha_i\}, \{\eta_{it}\}, \beta, \sigma_\eta^2, \sigma_\varepsilon^2, data\right)$

We sample from each distribution 1 through 5 in turn, after which we return back to step 1 and repeat. The resulting sequence of parameter draws forms a Markov chain, the stationary distribution of which is exactly the posterior distribution. Given a sample of draws of the posterior distribution, it is then straightforward to numerically integrate out the latent variables and obtain the marginal posterior of parameters, $f(\theta|data)$, or the distribution of the random effects, $f(\{\alpha_i\}|data)$ and $f(\{\eta_{it}\}|data)$, for example. We now discuss how to draw from each conditional distribution.

A1 Latent firm-year random effects

The firm-year random effects, η_{it}, are simulated using a Bayesian regression of the fund returns on a set of year indicator variables, with known variance. This is done on a firm-by-firm basis, as the random effects are assumed independent across firms (and time). For each firm, i, the regression model takes the form

$$y_i = X_i\beta + Z_i\eta_i + \varepsilon_i, \tag{A.5}$$

where y_i is a vector of stacked fund returns for the U_i funds of firm i, and X_i is the submatrix of the covariates $\left[X_{i1}' \dots X_{iU_i}'\right]'$ for which each row correspond to a fund of firm i. The vector η_i contains the firm-year random effects for the years in which firm i has at least one active fund. The length of the vector η_i is denoted T_i, and may vary by firm. The matrix Z_i is a $U_i \times T_i$ matrix of indicator variables. Each row represents a fund of firm i, and contains ones in the columns that correspond to the years that the fund is active, and zeros in all other columns.

Given the prior in equation (A.2), and using the standard Bayesian regression setup (e.g., Rossi, Allenby, and McCulloch [2005]), the posterior distribution is

$$\eta_i|\{\alpha_i\},\theta,data \sim \mathcal{N}\left(\mu_\eta,\sigma_\varepsilon^2\Omega^{-1}\right), \tag{A.6}$$

where

$$\Omega = \frac{\sigma_\varepsilon^2}{\sigma_\eta^2} \cdot \mathbb{I}_{T_i} + Z_i' Z_i \tag{A.7}$$

$$\mu_\eta = \Omega^{-1} \left(\alpha_i \cdot \frac{\sigma_\varepsilon^2}{\sigma_\eta^2} \cdot 1_{T_i} + Z_i'(y_i - X_i \beta) \right), \tag{A.8}$$

where $\mathbb{I}_{T_i}$ is the $T_i \times T_i$ identity matrix, and 1_{T_i} a $T_i \times 1$ vector of ones.

A2 Variance of fund-specific error term and β-coefficients

Given the conditioning on the random effects, η_{it}, this step is a standard Bayesian regression. With the conjugate prior

$$\sigma_\varepsilon^2 \sim IG(a_0, b_0) \tag{A.9}$$

$$\beta | \sigma_\varepsilon^2 \sim \mathcal{N}\left(\mu_0, \sigma_\varepsilon^2 \Sigma_0^{-1}\right), \tag{A.10}$$

the posterior distribution is

$$\sigma_\varepsilon^2 | \{\eta_i\}, data \sim IG(a, b) \tag{A.11}$$

$$\beta | \sigma_\varepsilon^2, \{\eta_i\}, data \sim \mathcal{N}\left(\mu, \sigma_\varepsilon^2 \Sigma^{-1}\right), \tag{A.12}$$

where

$$a = a_0 + \sum_{i=1}^{N} U_i \tag{A.13}$$

$$b = b_0 + e'e + (\mu - \mu_0) \Sigma_0 (\mu - \mu_0) \tag{A.14}$$

$$\Sigma = \Sigma_0 + X'X \tag{A.15}$$

$$\mu = \Sigma^{-1} \cdot \left(\Sigma_0 \mu_0 + X'(y - Z\eta)\right). \tag{A.16}$$

The vector $y = \left[y_i' \dots y_N'\right]'$ contains the fund returns stacked across the N firms, X is the matrix of stacked X_i and Z the stacked Z_i. The vector $e = y - Z\eta - X\mu$ contains the stacked error terms .

A3 Latent firm random effects

Drawing the firm random effects, α_i, is similar in spirit to simulating the firm-year random effects in step 1. Write the estimation problem as a regression of the firm-year random effects on a set of indicator variables

$$\eta = W\alpha + \nu, \tag{A.17}$$

where $\eta = [\eta_1 \dots \eta_N]'$, and $\alpha = [\alpha_1 \dots \alpha_N]'$, and $\nu \sim \mathcal{N}\left(0, \sigma_\eta^2 \cdot \mathbb{I}_N\right)$. The matrix W is a $\sum_{i=1}^N T_i \times N$ matrix of indicator variables. Each row of W represents a firm-year, and contains a one in the column of the corresponding firm, and zeros in all other columns.

With the prior in equation (A.3), the posterior distribution is

$$\alpha | \{\eta_{it}\}, \theta, data \sim \mathcal{N}\left(\mu_\alpha, \sigma_\eta^2 A^{-1}\right), \tag{A.18}$$

where

$$A = \frac{\sigma_\eta^2}{\sigma_\alpha^2} \cdot \mathbb{I}_N + W'W \tag{A.19}$$

$$\mu_\alpha = A^{-1}\left(W'\eta\right). \tag{A.20}$$

A4 Variance of firm-year random effects

The variance of the firm-year random effects, σ_η^2, is the variance of the residuals $v = \eta - W\alpha$ from the regression in step 3. Using the inverse gamma prior

$$\sigma_\eta^2 \sim IG(c_0, d_0), \tag{A.21}$$

yields the posterior distribution

$$\sigma_\eta^2 | \{\alpha_i\}, \{\eta_i\}, data \sim IG(c, d), \tag{A.22}$$

where

$$c = c_0 + \sum_{i=1}^{N} T_i \tag{A.23}$$
$$d = d_0 + v'v. \tag{A.24}$$

A5 Variance of firm random effects

The variance of the firm random effects, σ_α^2, using the inverse gamma prior

$$\sigma_\alpha^2 \sim IG(f_0, g_0), \tag{A.25}$$

has posterior distribution

$$\sigma_\alpha^2 | \{\alpha_i\}, data \sim IG(f, g), \tag{A.26}$$

with parameters:

$$f = f_0 + N, \tag{A.27}$$
$$g = g_0 + \alpha'\alpha. \tag{A.28}$$

A6 Mixture of Normals Specification

For the mixtures of Normals specification we replace the distribution of the fund-specific error term in equation (A.4), with a mixture of K Normal distributions,

$$\varepsilon_{iu} \sim \sum_{k=1}^{K} p_k \cdot \mathcal{N}\left(\mu_k, \sigma_{\varepsilon,k}^2\right). \tag{A.29}$$

Setting K=1 reduces the model to the baseline Normal specification in (A.4). We drop the intercept in X_{iu} because it is absorbed by the error term, which has mean $E[\varepsilon_{iu}] = \sum_{k=1}^{K} p_k \mu_k$. This specification is equivalent to the specification with an intercept in X_{iu} and zero mean error ε_{iu}, but it is easier to implement because it avoids enforcing cross-parameter restrictions on the μ_k. To estimate the mixture model by Gibbs sampler, the procedure requires one more latent variable that indicates which of the K Normal distributions each observation is drawn from. Conditional on this indicator, the Gibbs steps described above remain largely unchanged. For details on estimation of the latent indicator and the parameters of the mixture components we refer to West [1992], Diebolt and Robert [1994], and Chen and Liu [2000]. We use the algorithm proposed by Berkhof et al. [2003] and Marin and Robert [2008] to deal with the well-known label switching problem in calculating marginal likelihoods used to select the number of mixture components.

A7 Priors and Starting Values

Our Gibbs sampler uses 10,000 iterations for the initial burn-in, followed by 100,000 iterations to simulate the posterior distribution. During the burn-in phase, the simulations converge quickly. We use diffuse prior distributions for the parameters, so that our results are driven by the data rather than prior assumptions. First, we set $a_0 = 2.1$, and $b_0 = 1$. This implies that our prior belief is that $E[\sigma_\varepsilon] = 0.854$, and that σ_ε is between 0.362 and 2.874 with 99% probability (note that this is for ten-year fund returns, so the annualized volatility is about a factor 3 lower). Second, we set $c_0 = f_0 = 2.1$, and $d_0 = g_0 = 0.15^2$. Since both the $\alpha's$ and η's are specified at the annual level, this implies that $E[\sigma_\alpha] = E[\sigma_\eta] = 0.128$ per year, and σ_α and σ_η are between 0.054 and 0.431 (annually) with 99% probability. Conditional on X, the prior ten-year fund return variance, $100\sigma_\alpha^2 + 10\sigma_\eta^2 + \sigma_\varepsilon^2$, has an expected value of 1.658, and is between 0.861 and 4.666 with 99% probability. Finally, we set the prior mean for β equal to zero ($\mu_0 = 0$), implying a prior mean fund return of zero. We set Σ_0 equal to the identity matrix, so that the prior β's are between -3.1 and +3.1 with 99% probability.

For the mixtures of Normals specifications we set the prior of each mixture component, $1 \ldots K$, equal to the prior of the error term ε in the Normal model, i.e., mean zero and Inverse Gamma prior parameters equal to $a_0$and b_0. This ensures that the prior distribution of y is the same across all K, so that the Bayes Factor (see below) is a valid comparison across different mixtures. The prior distribution of the mixture probabilities, p, is the conjugate Dirichlet distribution, $Dir(K, \delta)$, with $\delta = 1_K \cdot 10$. This implies that all distributions in the mixture have equal prior mean probability, $1/K$.

We start the algorithm with all α's and β's equal to zero (their prior means). We initialize all variances (σ^2_α,σ^2_η, and σ^2_ε) at their prior means. For the mixtures specification, we set the mixture probabilities to their prior mean, $1/K$. We do not need starting values for the η's, since they are the first variables we simulate.

References

Albert, James and Siddhartha Chib, 1993, Bayesian Analysis of Binary and Polychotomous Response Data, *Journal of the American Statistical Association* 88, 669–679.

Bailey, Martin J., Richard F. Muth, and Hugh O. Nourse, 1963, A Regression Method for Real Estate Price Index Construction, *Journal of the American Statistical Association* 58, 933–942.

Baquero, Guillermo, Jenke ter ter Horst, and Marno Verbeek, 2005, Survival, Look-Ahead Bias, and Persistence in Hedge Fund Performance, *Journal of Financial and Quantitative Analysis* 40, 493–517.

Bengtsson, Ola, 2013, Relational Venture Capital Financing of Serial Founders, *Journal of Financial Intermediation* 3, 308–344.

Berkhof, Johannes, Iven van Mechelen and Andrew Gelman, 2003, A Bayesian Approach to the Selection and Testing of Mixture Models, *Statistica Sinica* 13, 423–442.

Billingsley, Patrick, 1961, Statistical Methods in Markov Chains, *The Annals of Mathematical Statistics* 32, 12–40.

Braun, Reiner, Tim Jenkinson, and Ingo Stoff, 2013, How Persistent is Private Equity Performance? Evidence from Deal-Level Data, *Working Paper*, Oxford University.

Burstein, L., K. H. Fischer, M. D. Miller, 1980, The Multilevel Effects of Background On Science Achievement: A Cross National Comparison, *Journal of Sociology* 53, 215–225.

Chen, Rong, and Jun S. Liu, 2000, Mixture Kalman Filters, *Journal of the Royal Statistical Society, Series B* 62, 493–508.

Chib, Siddhartha, 1995, Marginal Likelihood from the Gibbs Output, *Journal of the American Statistical Association* 90, 1313–1321.

Chung, Ji-Woong, 2010, Performance Persistence in Private Equity Funds, *Working Paper*, Chinese University of Hong Kong.

Cochrane, John, 1988, How Big is the Random Walk Component in GDP?, *Journal of Political Economy* 96, 893–920.

de Leeuw, Jan and Erik Meijer, 2008, Handbook of Multilevel Analysis, Springer, NY.

Diebolt, J., and Christian P. Robert, 1994, Estimation of Finite Mixture Distributions

Through Bayesian Sampling, *Journal of the Royal Statistical Society, Series B,* 56, 363–375.

Ewens, Michael, and Matthew Rhodes-Kropf, 2013, Is a VC Partnership Greater Than the Sum of its Partners, *Working Paper*, Harvard Business School.

Fisher, L., 1966, Some New Stock Market Indices, *Journal of Business* 39, 191–225.

Gelfand, Alan E., Sujit K. Sahu, and Bradley P. Carlin, 1995, Efficient parametrisations for normal linear mixed models, *Biometrika* 82, 479–488.

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

Harris, Robert, Tim Jenkinson, Steven Kaplan, 2013, Private Equity Performance: What Do We Know?, *Working Paper*, University of Chicago.

Harris, Robert, Tim Jenkinson, Steven Kaplan, and Ruediger Stucke, 2013, Has Persistence Persisted in Private Equity? Evidence From Buyout and Venture Capital Funds, *Working Paper*, University of Chicago.

Hochberg, Yael, Alexander Ljungqvist, and Anette Vissing-Jorgensen, 2014, Informational Hold-Up and Performance Persistence in Venture Capital, *Review of Financial Studies* 27, 102-152.

Jagannathan, Ravi, Alexey Malakhov, and Dmitry Novikov, 2010, Do Hot Hands Exist among Hedge Fund Managers? An Empirical Evaluation, *Journal of Finance* 65, 217–255.

Korteweg, Arthur G., 2013, Markov Chain Monte Carlo Methods in Corporate Finance, in P. Damien, P. Dellaportas, N. Polson, and D. Stephens, eds.: Bayesian Theory and Applications.

Korteweg, Arthur G., and Stefan Nagel, 2013, Risk-adjusting the Returns to Venture Capital, *Working Paper*, NBER working paper #1934.

Kaplan, Steven and Antoinette Schoar, 2005, Private Equity Performance: Returns, Persistence, and Capital Flows, *Journal of Finance* 60, 1791–1823.

Kass, Robert E., and Adrian E. Raftery, 1995, Bayes Factors, *Journal of the American Statistical Association* 90, 773–795.

Lerner, Josh, Antoinette Schoar, and Wan Wongsunwai, 2007, Smart Institutions, Foolish

Choices? The Limited Partner Performance Puzzle, *Journal of Finance* 62, 731–764.

Li, Yi, 2013, Reputation, Volatility and Performance Persistence of Private Equity, *Working Paper*, Princeton University.

Marin, Jean-Michel, and Christian P. Robert, 2008, Approximating the Marginal Likelihood in Mixture Models, *Working Paper*, INRIA and Université Paris-Dauphine.

Marquez, Robert, Vikram Nanda, Deniz Yavuz, 2011, Performance Persistence in Private Equity Fund Returns, *Working Paper*, Purdue.

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.

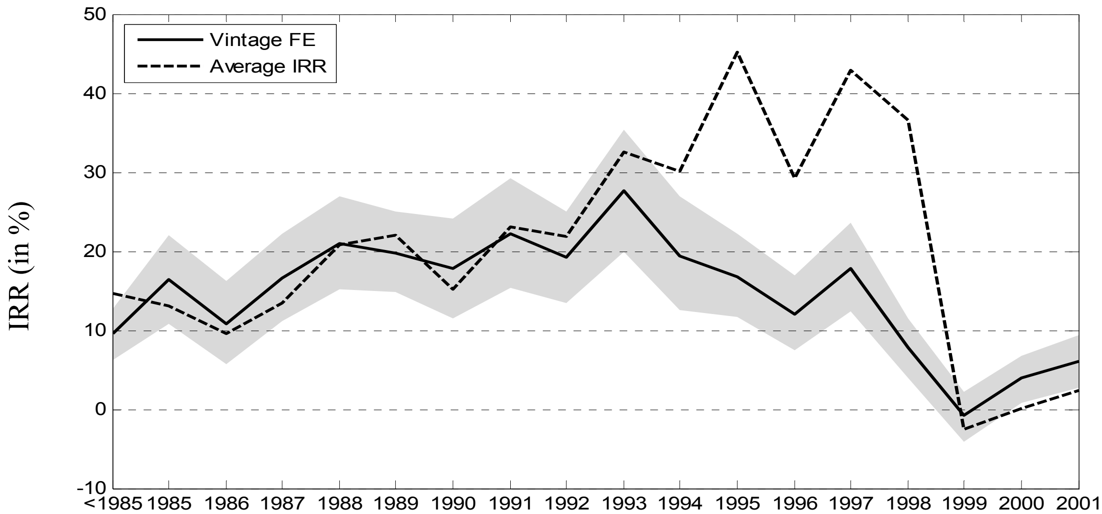
Rossi, Peter E., Greg M. Allenby and Robert McCullough, 2005, Bayesian Statistics and Marketing, John Wiley and Sons, Chichester, UK.

Sensoy, Berk, Yingdi Wang, and Michael Weisbach, 2013, Limited Partner Performance and the Maturing of the Private Equity Industry, *Working Paper*, NBER.

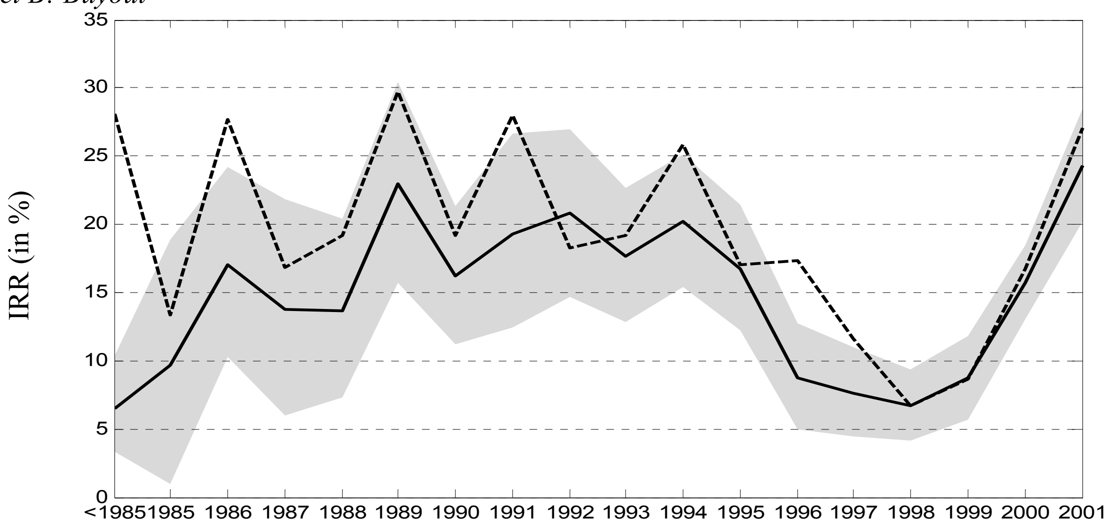
West, M., 1992, Modeling With Mixtures, in Bayesian Statistics 4, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Oxford University Press, 503–524.

Figure 1: Vintage Year Fixed Effects. Plots of the posterior mean of the vintage years fixed effects from Specification II in Table IV. The vintage year effects are for transformed to IRR equivalents (annualized, in percent). The shaded bands represents the (1%, 99%) Bayesian credible interval (confidence bounds). The broken line shows the average IRR calculated annually.

Panel A: VC



Panel B: Buyout



Panel C: Other

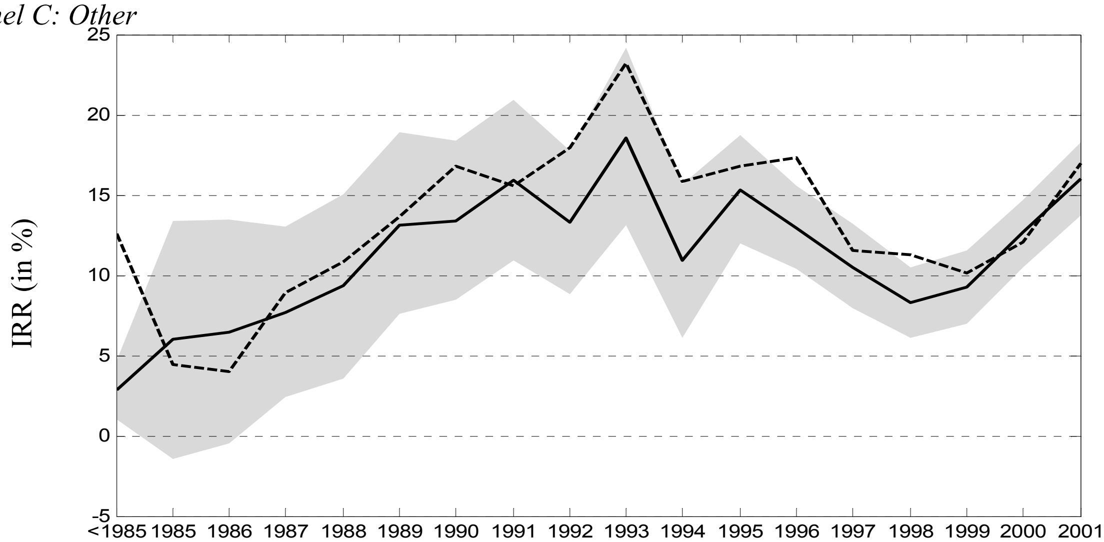


Figure 2: Fund Overlap and Covariance. The figure shows the covariance between total fund returns as a function of the overlap (in years) between two funds managed by the same firm, using the variance estimates for specification II in Table IV.

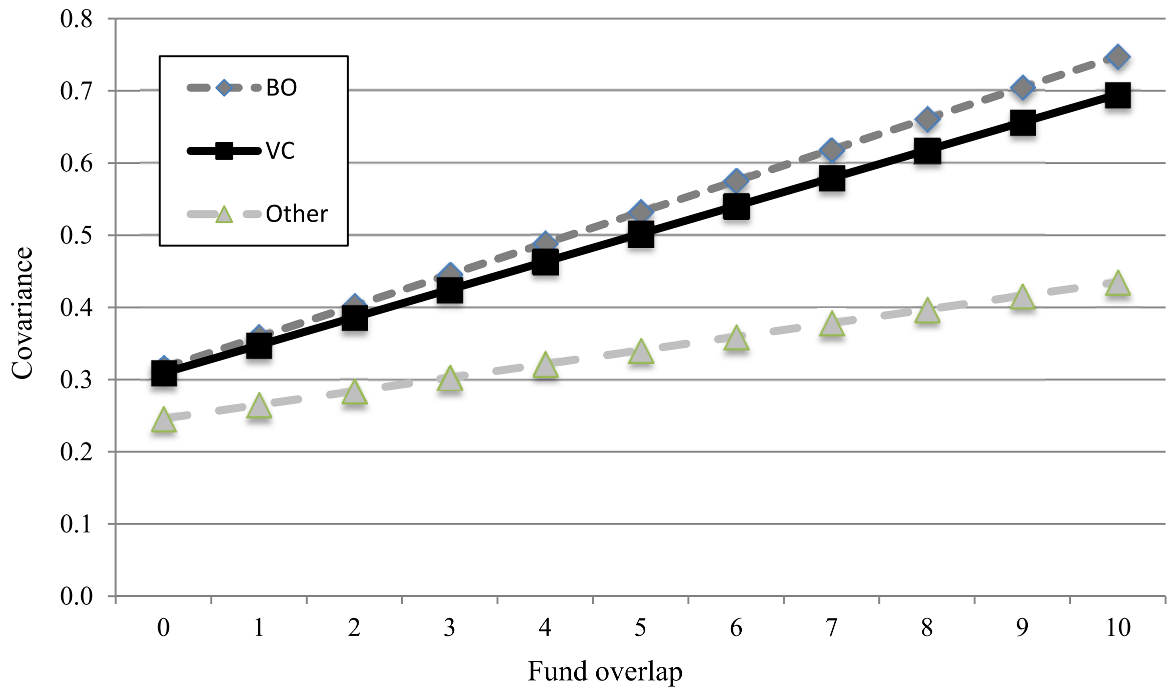


Figure 3: Estimates of Signal-to-Noise Ratio. Posterior distribution of the signal-to-noise ratio, s_α, by fund type, from the specifications reported in Table IV. The solid line is the kernel plot for specification I (without year fixed effects), and the broken line is the kernel plot for specification II in Table IV (with vintage year fixed effects).

Panel A: VC

Panel B: Buyout

Panel C: Other

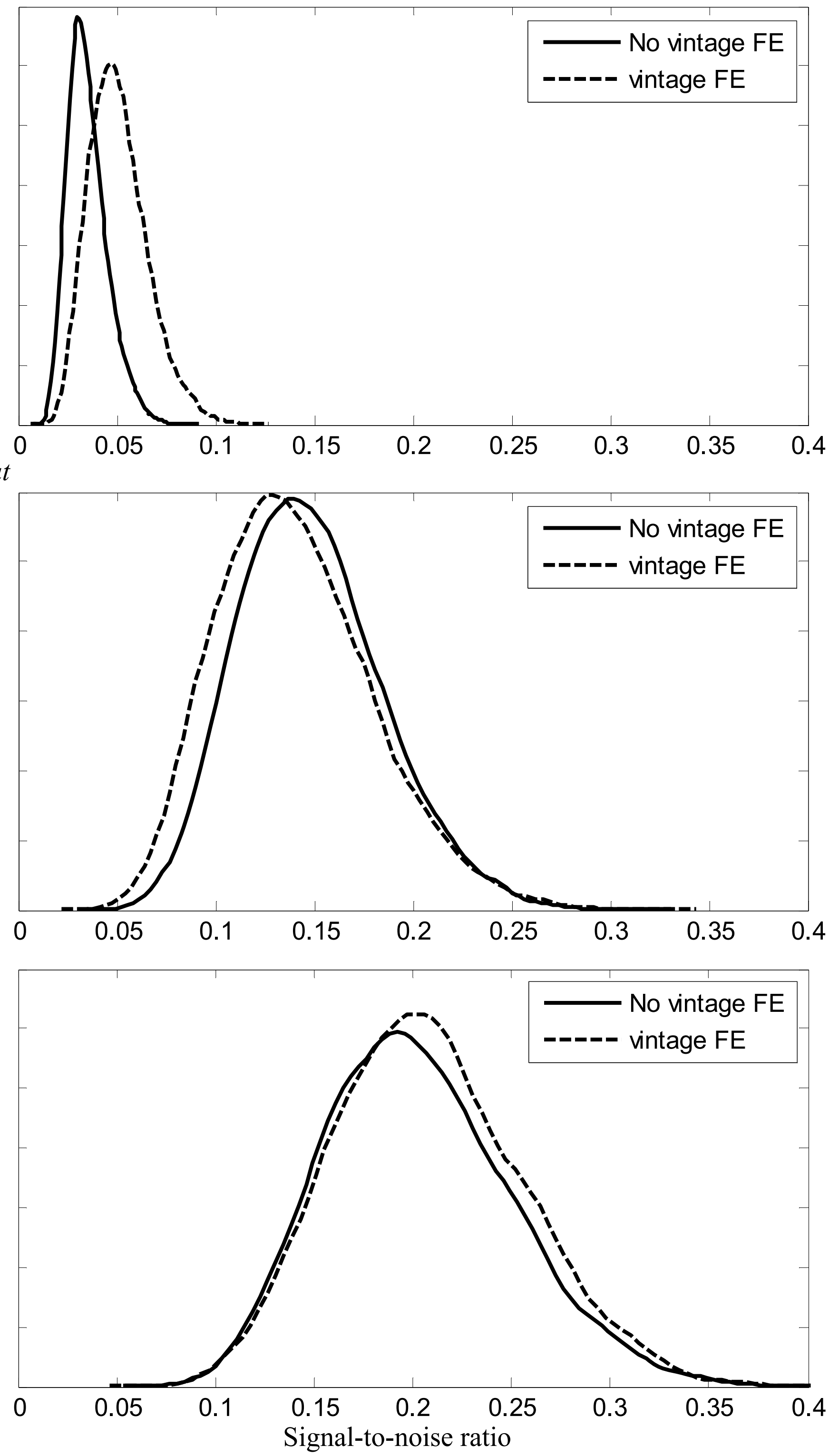


Figure 4: Speed of Learning. Graph of the posterior probability that a fund is in the top quartile of funds. Probabilities are calculated from 100,000 simulations of a panel of 100 firms, each with a different alpha that is drawn from the top 25% of the distribution. Each firm produces a sequence of 50 non-overlapping fund returns. Reported probabilities are averages of the posterior mean probability across the simulated firms after observing a given number of realized fund returns for each firm (*Fund history*, on the horizontal axis). The figure uses the parameter estimates from Table IV specification II, with vintage year fixed effects.

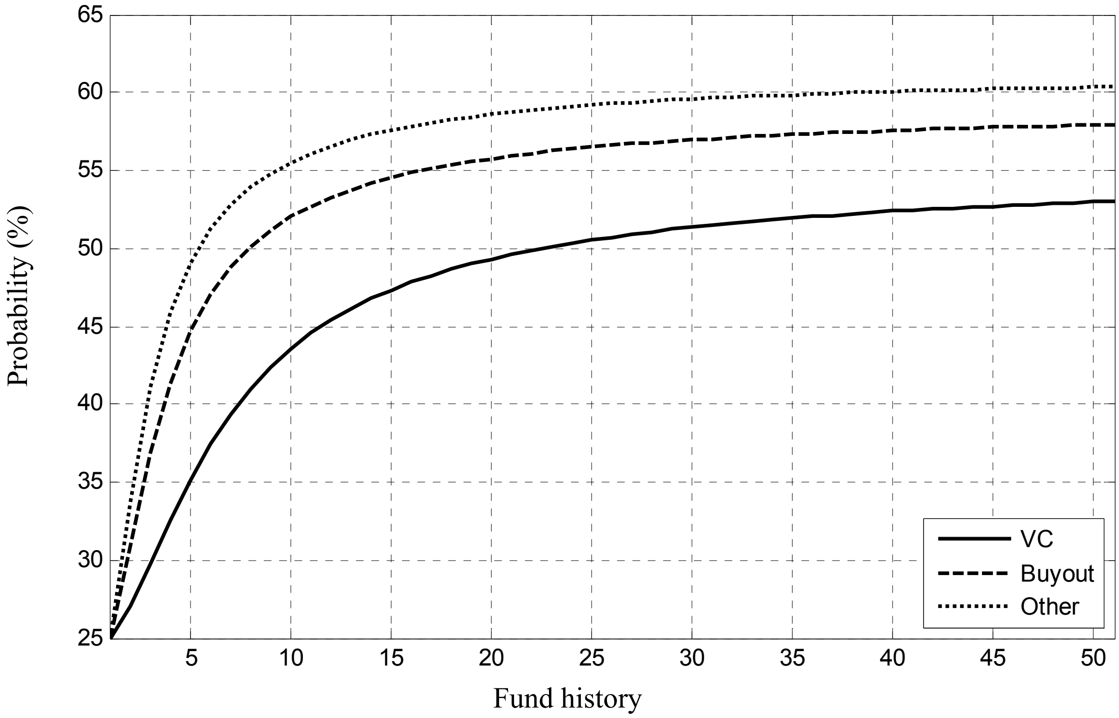


Figure 5: Investable Persistence. This figure shows the expected (true) alpha of investing in funds raised by PE firms with top-quartile performance as observed after a given number of realized fund returns for each firm (*Fund history*). Calculations are based on 100,000 simulations of a panel of fund histories for 100 firms, using the parameter estimates from Table IV specification II (with vintage year fixed effects).

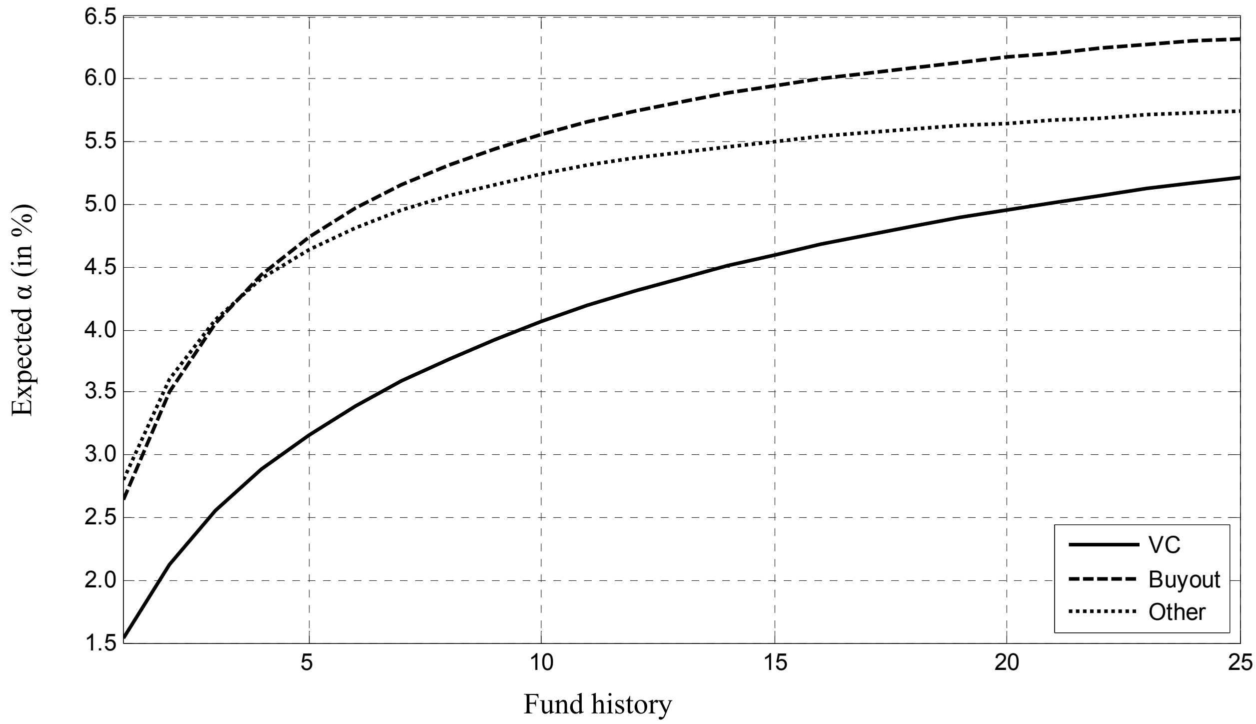
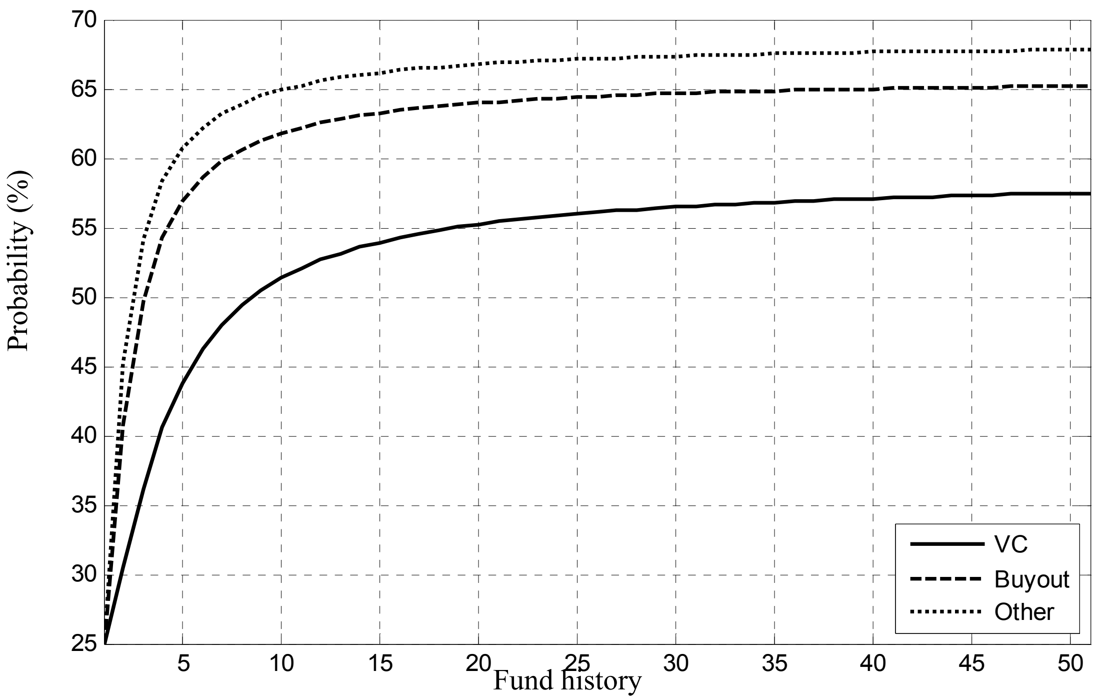


Figure 6: Speed of Learning in the Cash Flow Sample. Graph of the posterior probability that a fund is in the top quartile of funds, based on estimates for a subsample of funds for which cash flows between limited partners and the fund is available. Panels A and B show the probabilities when using IRR and PME as the performance measure, respectively. Probabilities are calculated as described in Figure 4, using the parameter estimates from Table VIII specification II (with vintage year fixed effects).

Panel A: IRR



Panel B: PME

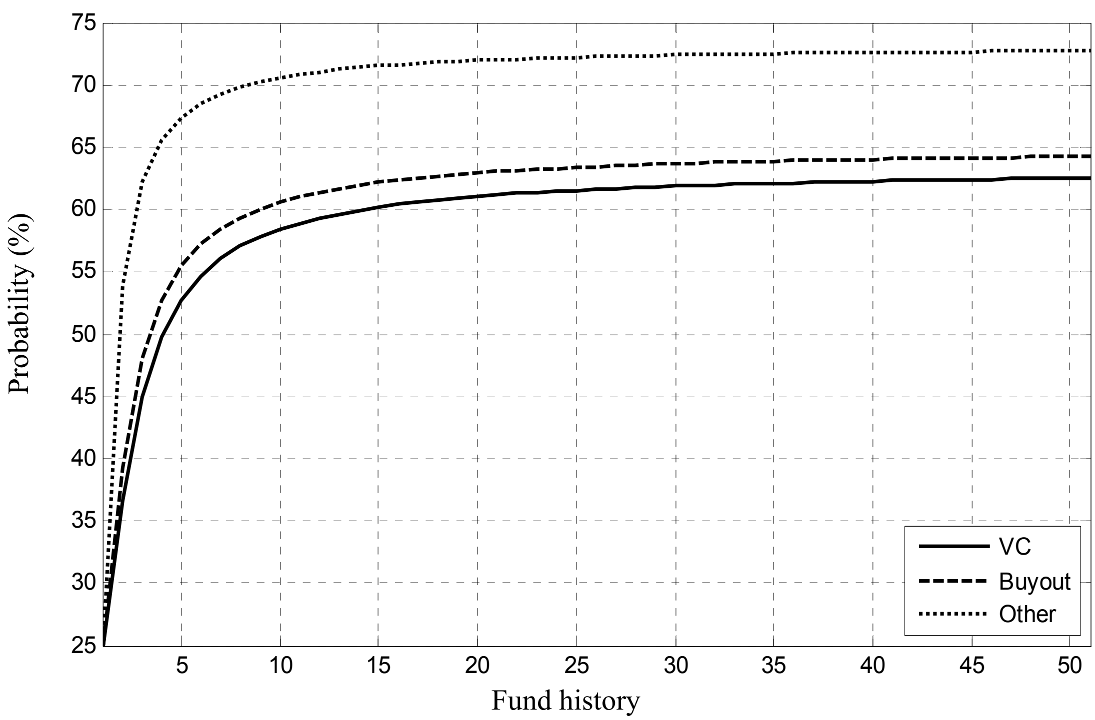
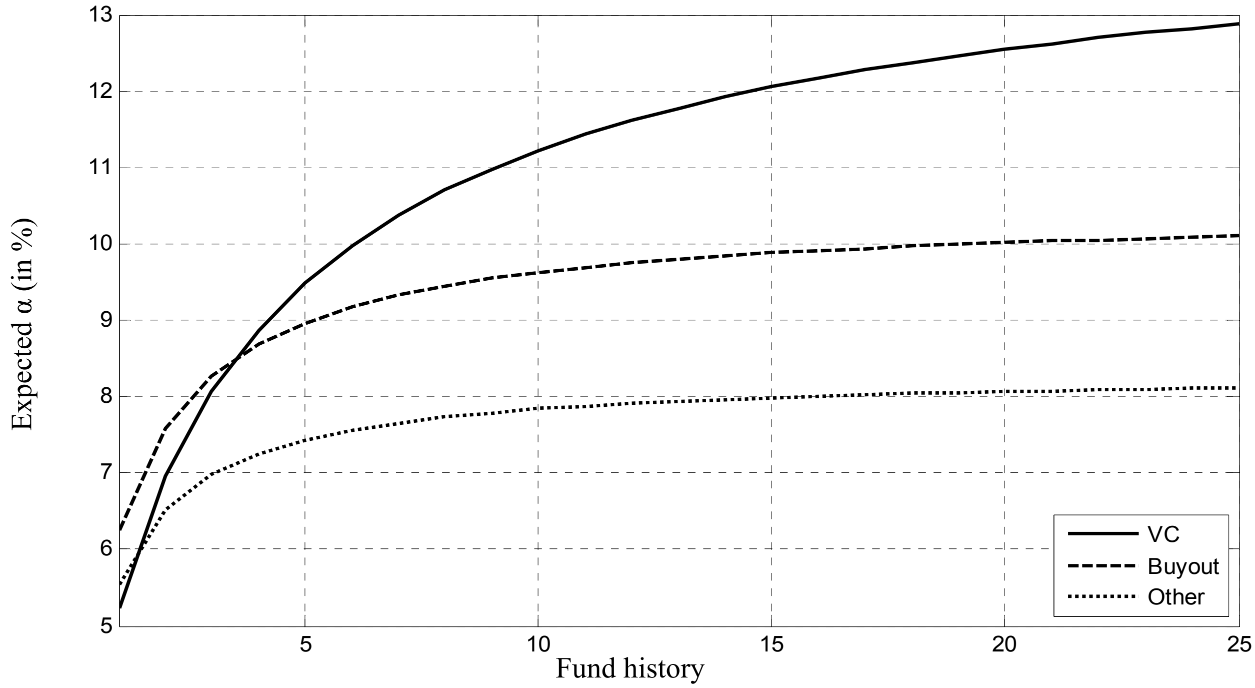


Figure 7: Investable Persistence in the Cash Flow Sample. This figure shows the expected (true) alpha of investing in funds raised by PE firms with top-quartile performance as observed after a given number of realized fund returns for each firm (*Fund history*). Calculations are based on 100,000 simulations of fund histories for 100 firms, using the parameter estimates from a subsample of funds with data on cash flows between limited partners and the fund (Table VIII specification II, which has vintage year fixed effects). Panel A shows the expected alpha when using IRR as the performance measure. Panel B show the alpha in the PME performance measure.

Panel A: IRR



Panel B: PME

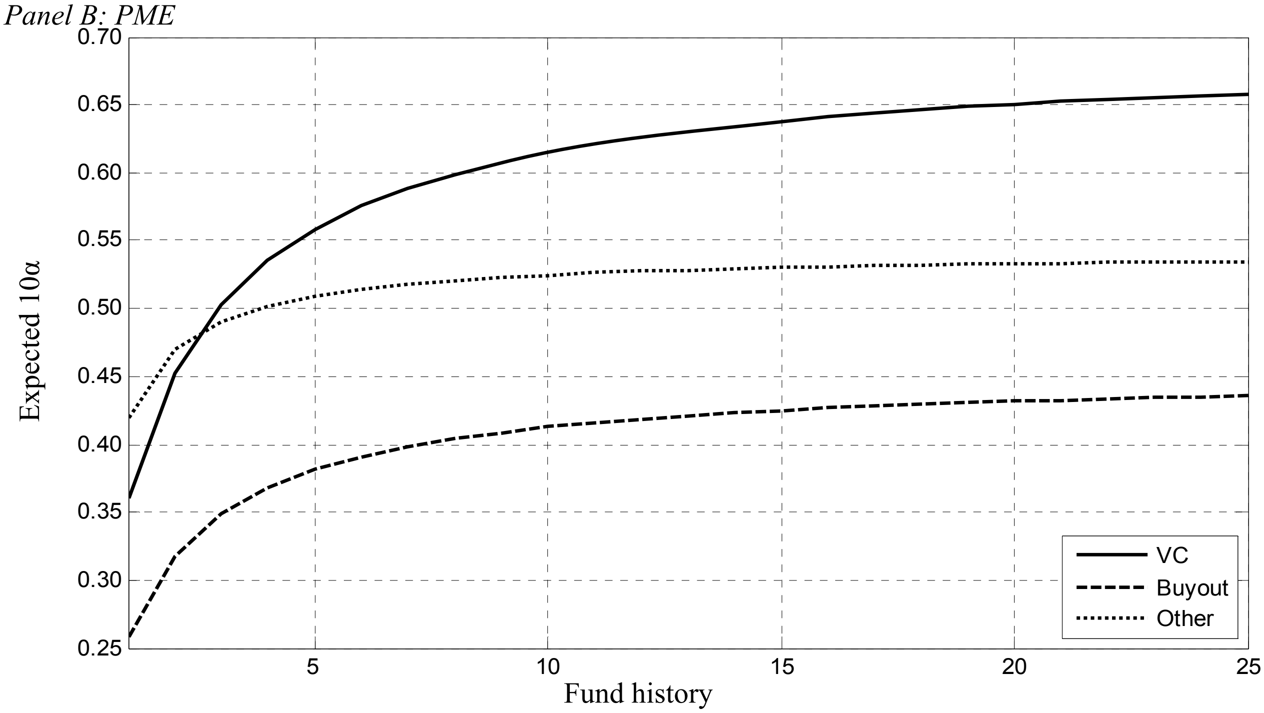
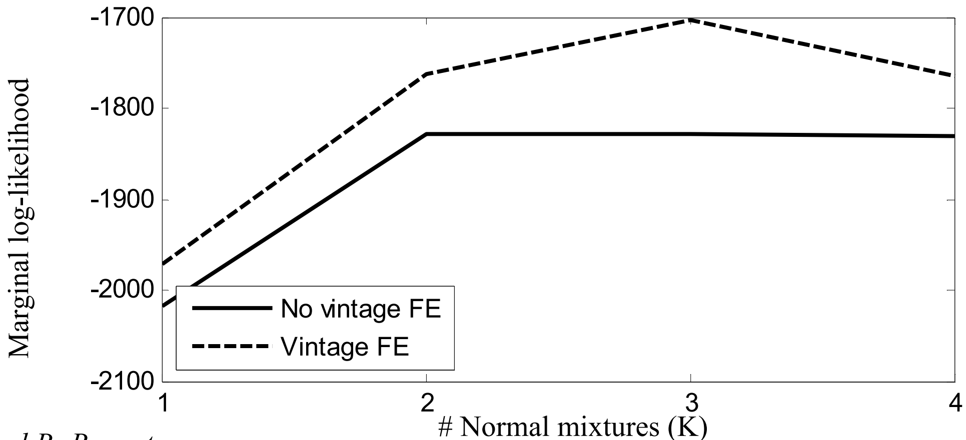
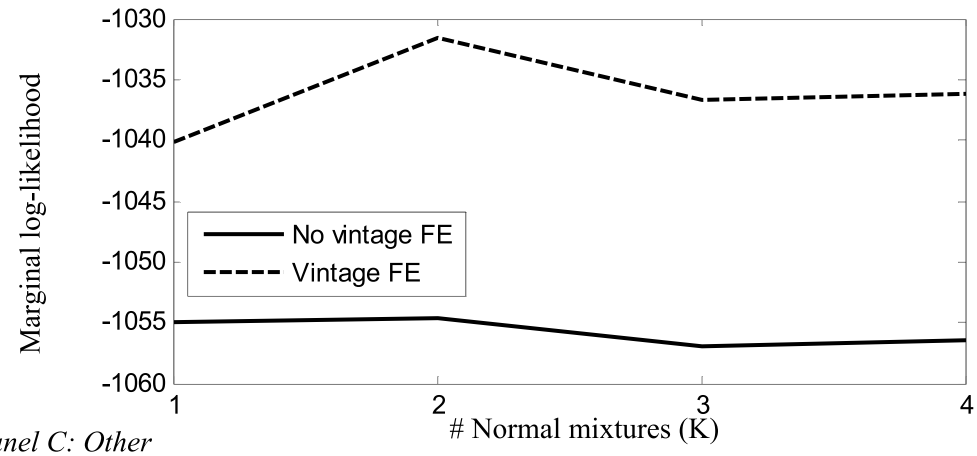


Figure A1: Marginal Log-likelihood. Plots of the marginal log-likelihood as a function of the number of Normal mixtures in the error term distribution, by fund type. The solid line represents specification I of Table IV (which has no vintage year fixed effects), and the striped line represents specification II (with vintage year fixed effects).

Panel A: VC



Panel B: Buyout



Panel C: Other

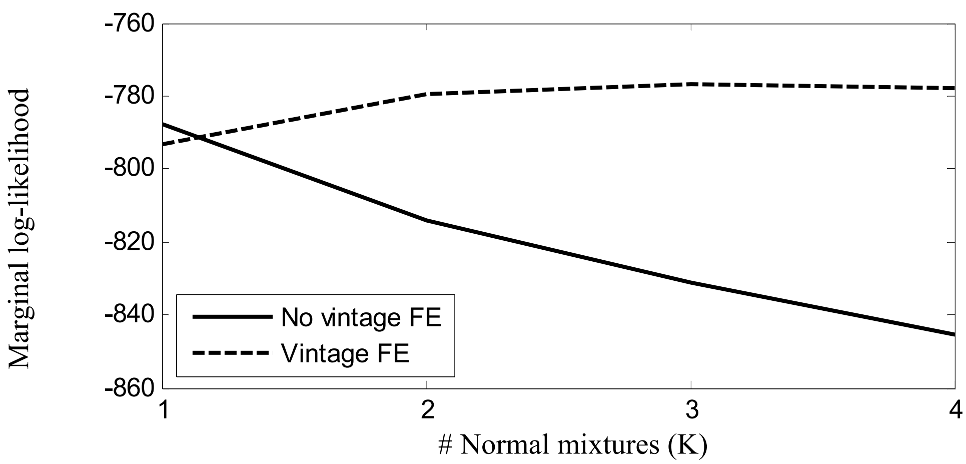


Table I: Summary Statistics. Descriptive statistics of the sample of private equity funds, by fund type (VC, Buyout, or Other). The sample contains 1,924 fully-liquidated funds raised between 1969 and 2001, with at least $5 million in committed capital (in 1990 dollars) and with non-missing returns data. The funds are raised by 831 individual PE firms (some firms manage funds of more than one fund type). Fund size is the committed capital in millions of dollars. *IRR* is the fund's internal rate of return, net of fees (where 0.1 represents a 10% return). The *Ten-year log return* is computed as $10 \cdot \ln(1 + IRR)$. *Overlap* is the number of years of overlap for funds of the same firm and type that overlap. Source: Preqin.

Panel A: Broad fund categories

	VC	Buyout	Other
# Funds	842	562	518
# Firms	409	285	197
# Funds / firm			
Mean	2.1	2.0	2.6
Median	1	1	2
Std. dev.	1.9	1.6	2.6
10^{th} percentile	1	1	1
90^{th} percentile	4	4	5
Fund size ($m)			
Mean	206.9	694.1	373.3
Median	110.0	300.0	206.8
Std. dev	276.1	1,035.6	517.1
10^{th} percentile	27.0	52.6	33.0
90^{th} percentile	500.0	1,823.6	863.0
IRR (in %)			
Mean	17.7	16.9	13.9
Median	8.6	14.9	11.9
Std. dev.	54.8	18.6	12.9
10^{th} percentile	-10.4	-1.7	0.4
90^{th} percentile	46.0	37.9	28.9
Ten-year log return			
Mean	1.173	1.438	1.245
Median	0.825	1.385	1.124
Std. dev.	2.623	1.552	1.075
10^{th} percentile	-1.101	-0.170	0.040
90^{th} percentile	3.786	3.216	2.542
Overlap (years)			
# fund pairs	891	512	968
Mean	5.8	5.8	6.8
Median	6	6	7
Std. dev.	2.5	2.3	2.4
10^{th} percentile	2	2	3
90^{th} percentile	9	9	9

Panel B: Fund sub-categories

		# funds	# firms	Fund size		IRR (in %)		
				mean	median	mean	median	std. dev.
VC								
	Early-stage	177	105	186.2	92.0	23.7	6.9	90.8
	Late-stage	153	89	289.0	154.0	12.3	10.5	17.0
	Generalist	512	239	187.5	104.0	17.2	8.5	44.7
Other								
	Real estate	202	86	389.3	263.0	13.8	12.9	9.7
	Fund-of-funds	144	48	349.3	138.3	11.2	8.3	13.7
	Distressed Debt	36	14	662.1	438.0	14.6	14.0	10.8
	Natural Resources	58	25	265.4	154.8	17.0	15.2	15.7
	Secondaries	34	12	336.8	263.5	16.8	15.0	12.2
	Infrastructure, Turnaround, Special Situations, Co-Investments, Venture Debt	44	26	293.0	150.0	16.9	13.0	18.6

Table II: Fund Internal Rates of Return by Vintage Year. This table shows the number of private equity funds in the sample, and their internal rates of return, by vintage year and fund type (VC, Buyout, and Other). The sample comprises 1,924 fully-liquidated funds over the period 1969 to 2001, with at least $5 million in committed capital (in 1990 dollars) and non-missing returns data. Weighted average IRRs are based on funds for which size data is available. Source: Preqin.

	VC				Buyout				Other			
		IRR (in %)				IRR (in %)				IRR (in %)		
Vintage year	Funds	Avg.	Median	Weighted Avg.	Funds	Avg.	Median	Weighted Avg.	Funds	Avg.	Median	Weighted Avg.
1969	1	8.7	8.7	8.7	0	-	-	-	0	-	-	-
1970	0	-	-	-	0	-	-	-	0	-	-	-
1971	0	-	-	-	0	-	-	-	0	-	-	-
1972	1	21.5	21.5	21.5	0	-	-	-	0	-	-	-
1973	0	-	-	-	0	-	-	-	0	-	-	-
1974	0	-	-	-	0	-	-	-	0	-	-	-
1975	0	-	-	-	0	-	-	-	0	-	-	-
1976	0	-	-	-	0	-	-	-	0	-	-	-
1977	0	-	-	-	1	35.5	35.5	35.5	0	-	-	-
1978	2	48.6	48.6	51.0	0	-	-	-	0	-	-	-
1979	1	18.5	18.5	18.5	1	19.4	19.4	19.4	1	18.0	18.0	18.0
1980	6	16.5	14.0	25.9	3	23.7	25.8	25.7	1	12.0	12.0	12.0
1981	7	18.9	11.3	19.1	0	-	-	-	2	11.1	11.1	-
1982	11	13.8	9.3	19.3	1	39.2	39.2	39.2	1	10.0	10.0	10.0
1983	12	9.8	9.3	9.4	2	21.9	21.9	21.4	3	14.7	5.9	36.5
1984	20	12.3	12.0	12.0	7	30.4	18.4	30.4	1	7.1	7.1	7.1
1985	21	13.1	13.0	14.9	4	13.4	10.7	17.9	4	4.4	2.7	15.3
1986	20	9.5	8.5	9.3	12	27.6	18.8	25.6	3	4.0	4.0	8.5
1987	22	13.4	14.8	12.6	9	16.8	18.9	9.7	8	8.9	8.3	10.7
1988	26	20.8	21.4	27.0	12	19.2	14.2	14.6	6	10.9	11.6	9.0
1989	35	22.0	16.4	30.1	12	29.8	27.5	28.6	8	13.6	11.1	14.7
1990	24	15.1	16.5	17.6	23	19.2	15.4	15.5	10	16.8	16.0	27.6
1991	18	23.1	21.1	27.2	10	28.0	25.2	28.4	10	15.5	12.2	18.3

1992	30	21.8	16.0	24.3	21	18.3	21.2	31.5	15	18.0	16.3	21.1
1993	36	32.6	29.5	35.9	23	19.2	16.9	19.7	16	23.2	19.8	24.6
1994	30	30.0	25.5	38.2	36	25.9	21.5	35.5	18	15.9	14.0	10.1
1995	41	45.2	17.5	42.5	33	17.1	17.6	15.8	30	16.8	17.0	17.6
1996	43	29.3	10.3	21.3	35	17.4	10.4	11.4	45	17.3	12.7	14.9
1997	62	42.8	20.6	37.6	54	11.6	9.0	9.3	45	11.5	8.4	10.8
1998	73	36.6	7.0	25.0	73	6.7	8.3	5.7	65	11.2	8.3	11.7
1999	95	-2.5	-3.5	-3.4	57	8.6	8.5	7.1	58	10.1	9.3	8.1
2000	120	0.1	-0.9	-0.2	90	16.8	17.5	16.3	80	12.1	12.4	13.4
2001	85	2.4	1.0	2.9	43	27.1	28.0	28.0	88	17.0	15.2	22.3

Table III: AR(1) Regressions. The table shows AR(1) regressions using fund IRRs, by fund type. The dependent variable is $IRR_{i,N}$, the IRR of fund *N* of PE firm *i*. In panel A, $IRR_{i,N-1}$ is the IRR of the most recent fund of the same type (VC, Buyout, Other) and the same firm, and $IRR_{i,N-2}$ is the IRR of the second previous fund. *VC=1* is a dummy variable that equals one for VC funds and zero otherwise. The variables *ln(Size)* and *ln(Sequence)* are the natural logarithm of fund's size and the sequence number of the fund within its class (VC, Buyout, Other) for a given PE firm. In Panel B, $IRR_{i,N-1}$ is the net IRR on the most recent, non-overlapping fund of the same type and the same firm. All regressions include vintage year fixed effects. Standard errors are clustered by firm, and are shown in brackets. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: All funds

	VC and Buyout					VC only		Buyout only		Other only	
$IRR_{i,N-1}$	0.162 *	0.304 ***			0.297 ***	0.129	0.270 ***	0.314 ***	0.299 ***	0.255 **	0.102
	(0.071)	(0.053)			(0.056)	(0.079)	(0.059)	(0.067)	(0.072)	(0.091)	(0.074)
$IRR_{i,N-2}$		-0.063	0.052		-0.064		-0.045		0.152		0.148
		(0.053)	(0.053)		(0.055)		(0.063)		(0.102)		(0.097)
$IRR_{i,N-3}$				-0.044							
				(0.026)							
VC=1	1.260	1.988	3.268	4.867	2.032						
	(2.696)	(4.024)	(4.714)	(6.237)	(4.835)						
ln(Size)					0.375						
					(1.649)						
ln(Sequence)					1.769						
					(4.506)						
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	673	375	398	244	365	411	240	262	135	295	167
Adj R^2	0.136	0.173	0.107	0.082	0.162	0.195	0.209	0.308	0.398	0.104	0.0692

Panel B: Non-overlapping funds

	VC and Buyout	VC	Buyout	Other
$IRR_{i,N-1}$	0.065	0.067	0.195	-0.296
	(0.090)	(0.067)	(0.181)	(0.163)
VC = 1	8.567			
	(7.901)			
Year FE	Y	Y	Y	Y
N	207	147	60	38
Adj R^2	0.064	0.087	-0.006	0.063

Table IV: Parameter Estimates. This table reports posterior means of parameters of the model described in the text. Posterior standard deviations (Bayesian standard errors) are in brackets. The error term ε_{iu} is a mixture of K normal distributions, where K is chosen as the best fit according to models' marginal log-likelihood. The model includes either a single intercept (specification I) or vintage year fixed effects, grouping the pre-1985 vintages into one bucket (specification II). The model is estimated separately for each fund type (VC, Buyout, Other), by Markov chain Monte Carlo (MCMC) using 10,000 burn-in cycles followed by 100,000 samples, saving every 10th draw. Panel A shows the parameter estimates, and panel B shows the variance decomposition estimates for the same parameters. Panel C shows the spread in alphas across percentiles of the posterior distribution.

Panel A: Parameter estimates

	VC		Buyout		Other	
	I	II	I	II	I	II
σ_α	0.049	0.055	0.060	0.056	0.049	0.049
	(0.007)	(0.008)	(0.008)	(0.008)	(0.006)	(0.006)
σ_η	0.258	0.193	0.142	0.203	0.202	0.135
	(0.031)	(0.037)	(0.039)	(0.043)	(0.028)	(0.028)
σ_ε	2.449	2.326	1.359	1.225	0.807	0.865
	(0.123)	(0.121)	(0.058)	(0.073)	(0.039)	(0.050)
Vintage FE	N	Y	N	Y	N	Y
K	3	3	2	2	1	3
N	842	842	562	562	518	518

Panel B: Variance decomposition

	VC		Buyout		Other	
	I	II	I	II	I	II
$100\cdot\sigma_\alpha^2$	0.243	0.309	0.361	0.316	0.244	0.246
	(0.067)	(0.087)	(0.094)	(0.089)	(0.065)	(0.061)
$10\cdot\sigma_\eta^2$	0.675	0.386	0.216	0.432	0.416	0.189
	(0.158)	(0.141)	(0.113)	(0.160)	(0.111)	(0.076)
σ_ε^2	6.015	5.426	1.852	1.505	0.654	0.751
	(0.604)	(0.567)	(0.159)	(0.180)	(0.064)	(0.087)
σ_y^2	6.933	6.120	2.428	2.253	1.314	1.186
	(0.596)	(0.561)	(0.152)	(0.168)	(0.084)	(0.090)
Signal-to-noise	0.035	0.051	0.148	0.141	0.185	0.208
	(0.010)	(0.015)	(0.037)	(0.040)	(0.045)	(0.048)

Panel C: Alpha spread

	VC		Buyout		Other	
	I	II	I	II	I	II
$q_\alpha(75\%)$-$q_\alpha(25\%)$	6.59%	7.42%	8.03%	7.51%	6.61%	6.64%
	(0.90%)	(1.05%)	(1.05%)	(1.06%)	(0.87%)	(0.81%)

$q_\alpha(87.5\%)-q_\alpha(12.5\%)$	11.24%	12.66%	13.70%	12.81%	11.27%	11.34%
	(1.54%)	(1.78%)	(1.79%)	(1.81%)	(1.49%)	(1.40%)

Table V: Speed of Learning. This table shows the expected alpha of investing in the top quartile of a universe of 100 PE firms, after observing a history of non-overlapping funds of a given length (*Fund history*). Expected alphas are calculated from 100,000 simulations of 100 firms, and reported after observing a sequence of 1 through 25 funds for each firm. The first four columns use the parameter estimates from Table IV specification I (without vintage year fixed effects), and the last four columns show specification II (with vintage year fixed effects). The table reports the expected true α of investing in the top quartile of firms, the probability of achieving a true alpha higher than zero, the probability of attaining a higher true alpha compared to a strategy of random allocations to funds, and the expected true α from a strategy of going long the top quartile and short the bottom quartile of funds.

Panel A: VC

	Without vintage year FE				With vintage year FE			
Fund history	E[α] top quartile	Prob(α>0)	Prob(α > random)	E[α] top -bottom quartile	E[α] top quartile	Prob(α>0)	Prob(α > random)	E[α] top -bottom quartile
1	1.15%	88.1%	82.0%	2.30%	1.54%	92.1%	86.4%	3.08%
2	1.60%	95.1%	90.1%	3.20%	2.13%	97.6%	93.8%	4.26%
3	1.93%	97.8%	92.7%	3.86%	2.55%	99.2%	95.8%	5.10%
4	2.19%	99.0%	96.3%	4.38%	2.88%	99.7%	98.3%	5.77%
5	2.41%	99.5%	97.4%	4.83%	3.16%	99.9%	98.9%	6.31%
10	3.18%	100.0%	100.0%	6.36%	4.06%	100.0%	100.0%	8.12%
25	4.25%	100.0%	100.0%	8.50%	5.22%	100.0%	100.0%	10.43%

Panel B: Buyout

	Without vintage year FE				With vintage year FE			
Fund history	E[α] top quartile	Prob(α>0)	Prob(α > random)	E[α] top -bottom quartile	E[α] top quartile	Prob(α>0)	Prob(α > random)	E[α] top -bottom quartile
1	2.92%	99.5%	97.5%	5.84%	2.65%	99.3%	97.1%	5.29%
2	3.85%	100.0%	99.6%	7.71%	3.50%	100.0%	99.5%	7.01%
3	4.44%	100.0%	99.8%	8.88%	4.05%	100.0%	99.8%	8.10%
4	4.86%	100.0%	100.0%	9.72%	4.44%	100.0%	100.0%	8.88%
5	5.17%	100.0%	100.0%	10.34%	4.74%	100.0%	100.0%	9.47%
10	6.04%	100.0%	100.0%	12.07%	5.56%	100.0%	100.0%	11.12%
25	6.82%	100.0%	100.0%	13.64%	6.33%	100.0%	100.0%	12.65%

Panel C: Other

	Without vintage year FE				With vintage year FE			
Fund history	E[α] top quartile	Prob(α>0)	Prob(α > random)	E[α] top -bottom quartile	E[α] top quartile	Prob(α>0)	Prob(α > random)	E[α] top -bottom quartile
1	2.88%	99.9%	99.0%	5.76%	2.80%	99.9%	99.0%	5.59%
2	3.72%	100.0%	99.9%	7.44%	3.60%	100.0%	99.9%	7.20%
3	4.21%	100.0%	100.0%	8.43%	4.08%	100.0%	100.0%	8.16%
4	4.55%	100.0%	100.0%	9.10%	4.40%	100.0%	100.0%	8.80%

5	4.80%	100.0%	100.0%	9.59%	4.64%	100.0%	100.0%	9.27%
10	5.44%	100.0%	100.0%	10.87%	5.24%	100.0%	100.0%	10.48%
25	5.97%	100.0%	100.0%	11.94%	5.74%	100.0%	100.0%	11.49%

Table VI: Sub-samples. This table shows estimates of signal-to-noise ratio, s_α, for different subsamples. The columns labeled *mean* and *std. dev.* report the posterior mean and standard deviation of the variance decomposition, respectively. *IQR* is the interquartile range of the posterior distribution of s_α, i.e., the spread in between the 75th and 25th percentile of the posterior distribution. *N* is the number of funds in the subsample. The estimates are based on model specification II in Table IV (which includes vintage year fixed effects). The top row corresponds to the signal-to-noise ratio of the full sample, as depicted in Figure 4. The number of funds in the fund size sub-samples do not add up to the number of funds in the full sample, because size is not observed for some funds.

Panel A: Long-term persistence

	VC				Buyout				Other			
	$100\sigma^2_\alpha$	$10\sigma^2_\eta$	σ^2_y	$q_\alpha(75\%)$-$q_\alpha(25\%)$	$100\sigma^2_\alpha$	$10\sigma^2_\eta$	σ^2_y	$q_\alpha(75\%)$-$q_\alpha(25\%)$	$100\sigma^2_\alpha$	$10\sigma^2_\eta$	σ^2_y	$q_\alpha(75\%)$-$q_\alpha(25\%)$
Full sample	0.309	0.386	6.120	7.42%	0.316	0.432	2.253	7.51%	0.246	0.189	1.186	6.64%
Fund size												
Small (< median)	0.823	0.522	7.537	12.00%	0.489	0.207	2.667	9.25%	0.326	0.245	1.110	7.54%
Large (>= median)	0.239	0.328	4.808	6.46%	0.234	0.452	1.107	6.40%	0.203	0.086	1.217	5.98%
GP location												
US	0.300	0.339	5.897	7.26%	0.286	0.438	2.046	6.99%	0.089	0.319	1.130	3.89%
Europe (incl. UK)	0.491	0.151	2.743	9.19%	0.427	0.146	1.578	8.68%	0.327	0.121	1.031	7.49%
ROW	0.562	0.258	10.465	9.72%	0.444	0.152	5.643	8.56%	1.434	0.180	2.306	15.45%
Style												
Early-stage	0.217	0.106	9.866	6.13%	-	-	-	-	-	-	-	-
Late-stage	0.272	0.272	1.949	6.87%	-	-	-	-	-	-	-	-
Generalist	0.377	0.386	5.674	8.10%	-	-	-	-	-	-	-	-
Real Estate	-	-	-	-	-	-	-	-	0.149	0.079	0.674	5.16%
Fund-of-Funds	-	-	-	-	-	-	-	-	0.182	0.074	1.056	5.65%
Sample period												
Early (< 1997)	0.529	0.429	4.947	9.68%	0.611	0.372	2.879	10.26%	0.338	0.123	1.239	7.71%
Late (>= 1997)	0.288	0.117	7.383	7.11%	0.303	0.156	1.709	7.33%	0.279	0.097	1.143	7.04%

Panel B: Signal-to-noise ratio

	VC				Buyout				Other			
	mean	std. dev.	IQR	N	mean	std. dev.	IQR	N	mean	std. dev.	IQR	N
Full sample	0.051	0.015	0.020	842	0.141	0.040	0.053	562	0.208	0.048	0.065	518
Fund size												
Small (< median)	0.111	0.043	0.057	373	0.183	0.068	0.096	261	0.291	0.103	0.146	210
Large (>= median)	0.051	0.022	0.028	374	0.123	0.049	0.064	270	0.168	0.056	0.077	210
GP location												
US	0.051	0.020	0.027	675	0.140	0.065	0.092	416	0.079	0.042	0.050	439
Europe (incl. UK)	0.180	0.079	0.110	93	0.269	0.081	0.112	113	0.310	0.116	0.159	59
ROW	0.057	0.036	0.042	74	0.083	0.058	0.061	33	0.584	0.187	0.292	20
Style												
Early-stage	0.023	0.011	0.013	177	-	-	-	-	-	-	-	-
Late-stage	0.141	0.063	0.080	153	-	-	-	-	-	-	-	-
Generalist	0.067	0.029	0.040	512	-	-	-	-	-	-	-	-
Real Estate	-	-	-	-	-	-	-	-	0.222	0.061	0.084	202
Fund-of-Funds	-	-	-	-	-	-	-	-	0.173	0.061	0.080	144
Sample period												
Early (< 1997)	0.108	0.035	0.047	407	0.211	0.091	0.129	245	0.271	0.088	0.125	227
Late (>= 1997)	0.040	0.016	0.021	435	0.178	0.058	0.080	317	0.243	0.066	0.091	291

Table VII: Summary Statistics for Cash Flow Sample. This table reports descriptive statistics of the "cash flow sample", a subsample of nearly-liquidated funds in Preqin with vintages between 1979 and 2001 for which reported cash flows between the fund and limited partners are reported. Nearly-liquidated funds are defined as funds with net asset value (NAV) less than or equal to 5% of the fund's commitment size by the end of the sample (in early 2013), while at least 50% of total commitment was called at some point in their history. *TVPI* is the total value to paid-in capital performance measure, calculated as the sum of cash distributions and any remaining NAV, divided by the sum of capital calls. *PME* is the public market equivalent measure of Kaplan and Schoar (2005), and is computed as the sum of discounted distributions and final NAV, divided by the sum of discounted capital calls. The discount rate is the realized market return. The table reports means, with medians in brackets. For comparison, the table also shows statistics for the "broad sample" of funds used in prior tables. These funds report TVPIs and IRRs but not cash flows. The t-statistics (in the columns labeled *t-stat*) are for differences in means test between the cash flow and broad samples. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	VC			BO			Other		
	Cash flow sample	Broad sample	t-stat	Cash flow sample	Broad sample	t-stat	Cash flow sample	Broad sample	t-stat
# Funds	187	842		181	562		72	518	
# Firms	118	409		119	285		46	197	
# Funds / firm	1.6	2.1	-2.554 **	1.5	2.0	-2.930 ***	1.6	2.6	-2.731 ***
	(1.0)	(1.0)		(1.0)	(1.0)		(1.0)	(2.0)	
Fund size ($m)	177.6	206.9	-1.328	625.5	694.1	-0.784	680.0	373.3	4.207 ***
	(130.0)	(110.0)		(309.3)	(300.0)		(425.0)	(206.8)	
Vintage year	1993.9	1994.7	-1.728 *	1994.5	1995.8	-3.154 ***	1996.2	1996.9	-1.298
	(1995.0)	(1997.0)		(1996.0)	(1997.0)		(1997.0)	(1998.0)	
TVPI	2.4	2.1	1.232	1.9	2.0	-1.290	1.7	1.7	-0.241
	(1.5)	(1.5)		(1.7)	(1.8)		(1.5)	(1.5)	
IRR (in %)	20.8	17.7	0.702	13.0	16.9	-2.536 **	13.2	13.9	-0.485
	(10.3)	(8.6)		(12.7)	(14.9)		(12.9)	(11.9)	
PME	1.4	-	-	1.2	-	-	1.3	-	-
	(0.9)	-		(1.2)	-		(1.2)	-	

Table VIII: Parameter Estimates for Cash Flow Sample. This table reports posterior means of parameters of the model described in the text, using a subsample of funds for which cash flows between the fund and limited partners is available. Posterior standard deviations (Bayesian standard errors) are in brackets. The dependent variable in panel A is the natural logarithm of the 10-year IRR. In panel B, the dependent variable is the natural logarithm of the public market equivalent (PME) measure of Kaplan and Schoar (2005). Explain alpha spread in panel B: this is the spread in log-PME (based on 10*alpha). See Table IV for description of the reported parameters and variables. For ease of interpretation, the alpha spreads in panel B are computed for 10α, as described in the text.

Panel A: Dependent variable = 10 ln(1+IRR)

	VC		Buyout		Other	
	I	II	I	II	I	II
$100\cdot\sigma_\alpha^2$	1.057	1.374	0.814	0.706	0.451	0.448
	(0.470)	(0.552)	(0.299)	(0.232)	(0.161)	(0.154)
$10\cdot\sigma_\eta^2$	0.306	0.206	0.527	0.162	0.132	0.132
	(0.404)	(0.216)	(0.609)	(0.139)	(0.111)	(0.099)
σ_ε^2	10.454	8.723	1.119	1.116	0.660	0.436
	(1.590)	(1.565)	(0.379)	(0.231)	(0.146)	(0.099)
σ_y^2	11.817	10.302	2.461	1.985	1.243	1.016
	(1.497)	(1.470)	(0.317)	(0.230)	(0.182)	(0.161)
Signal-to-noise	0.091	0.136	0.332	0.354	0.358	0.435
	(0.042)	(0.058)	(0.113)	(0.101)	(0.098)	(0.105)
$q_\alpha(75\%)-q_\alpha(25\%)$	13.55%	15.49%	11.96%	11.18%	8.92%	8.90%
	(2.94%)	(3.17%)	(2.26%)	(1.86%)	(1.58%)	(1.52%)
$q_\alpha(87.5\%)-q_\alpha(12.5\%)$	23.11%	26.42%	20.40%	19.07%	15.21%	15.18%
	(5.02%)	(5.41%)	(3.85%)	(3.17%)	(2.70%)	(2.59%)
Vintage FE	N	Y	N	Y	N	Y
K	2	3	3	3	1	1
N	187	187	181	181	72	72

Panel B: Dependent variable = ln(PME)

	VC		Buyout		Other	
	I	II	I	II	I	II
$100\cdot\sigma_\alpha^2$	0.308	0.308	0.147	0.134	0.206	0.186
	(0.104)	(0.094)	(0.031)	(0.030)	(0.050)	(0.044)
$10\cdot\sigma_\eta^2$	0.093	0.066	0.042	0.038	0.057	0.049
	(0.056)	(0.035)	(0.017)	(0.014)	(0.027)	(0.021)
σ_ε^2	0.835	0.762	0.259	0.235	0.111	0.076
	(0.171)	(0.156)	(0.062)	(0.051)	(0.026)	(0.017)
σ_y^2	1.236	1.136	0.448	0.407	0.373	0.311

	(0.154)	(0.148)	(0.062)	(0.052)	(0.057)	(0.048)
Signal-to-noise	0.251	0.273	0.331	0.331	0.548	0.595
	(0.082)	(0.080)	(0.071)	(0.070)	(0.078)	(0.075)
$q_{10\alpha}(75\%)-q_{10\alpha}(25\%)$	0.738	0.741	0.514	0.491	0.608	0.579
	(0.123)	(0.112)	(0.054)	(0.053)	(0.073)	(0.067)
$q_{10\alpha}(87.5\%)-q_{10\alpha}(12.5\%)$	1.259	1.264	0.877	0.837	1.037	0.987
	(0.210)	(0.190)	(0.091)	(0.091)	(0.125)	(0.115)
Vintage FE	N	Y	N	Y	N	Y
K	3	3	2	2	1	1
N	187	187	181	181	72	72

Table A1: Model Specification Tests. This table shows tests of the model specification. Column I reproduces specification I of Table IV. Column II drops the transient firm effect, η, from the model, and Column III drops the long-run firm-specific effect, α. Columns IV to VI show the same for specification II of Table IV, which includes vintage year fixed effects. The *Bayes factor* represents the ratio of marginal likelihoods, indicating the weight of evidence of each model relative to the full model specification in column IV, where a Bayes Factor of one indicates that the two models have equal support in the data. For each model the number of distributions in the error term (*K*) is chosen to find the best model fit by marginal log-likelihood. Posterior standard deviations (Bayesian standard errors) are in brackets.

Panel A: VC

	I	II	III	IV	V	VI
σ_α	0.049	0.042		0.055	0.040	
	(0.007)	(0.004)		(0.008)	(0.004)	
σ_η	0.258		0.313	0.193		0.268
	(0.031)		(0.023)	(0.037)		(0.025)
σ_ε	2.449	2.604	2.431	2.326	2.465	2.316
	(0.123)	(0.107)	(0.114)	(0.121)	(0.115)	(0.123)
σ_y^2	6.933	6.971	6.906	6.120	6.255	6.106
	(0.596)	(0.562)	(0.553)	(0.561)	(0.572)	(0.567)
Vintage FE	N	N	N	Y	Y	Y
K	3	3	2	3	3	3
N	842	842	842	842	842	842
Marginal log-L	-1,829.5	-1,755.4	-1,823.6	-1,703.8	-1,669.2	-1,727.4
Bayes factor	0.000	0.000	0.000	N/A	1.0E+15	0.000

Panel B: Buyout

	I	II	III	IV	V	VI
σ_α	0.060	0.043		0.056	0.044	
	(0.008)	(0.005)		(0.008)	(0.005)	
σ_η	0.142		0.277	0.203		0.279
	(0.039)		(0.034)	(0.043)		(0.020)
σ_ε	1.359	1.542	1.261	1.225	1.469	1.196
	(0.058)	(0.051)	(0.065)	(0.073)	(0.054)	(0.071)
σ_y^2	2.428	2.567	2.373	2.253	2.359	2.217
	(0.152)	(0.163)	(0.121)	(0.168)	(0.163)	(0.167)
Vintage FE	N	N	N	Y	Y	Y
K	2	2	1	2	2	2
N	562	562	562	562	562	562
Marginal log-L	-1,054.6	-1,039.7	-1,054.6	-1,031.6	-1,019.7	-1,035.0
Bayes factor	0.000	0.000	0.000	N/A	1.4E+05	0.033

Panel C: Other

	I	II	III	IV	V	VI
σ_α	0.049	0.039		0.049	0.037	
	(0.006)	(0.004)		(0.006)	(0.004)	
σ_η	0.202		0.245	0.135		0.199
	(0.028)		(0.022)	(0.028)		(0.017)
σ_ε	0.807	1.059	0.793	0.865	1.038	0.848
	(0.039)	(0.033)	(0.041)	(0.050)	(0.037)	(0.050)
σ_y^2	1.314	1.272	1.234	1.186	1.217	1.119
	(0.084)	(0.075)	(0.079)	(0.090)	(0.081)	(0.087)
Vintage FE	N	N	N	Y	Y	Y
K	1	2	1	3	3	3
N	518	518	518	518	518	518
Marginal log-L	-787.8	-777.6	-781.7	-776.9	-771.3	-774.5
Bayes factor	0.000	0.491	0.008	N/A	256.185	10.677