

Picking winners

Oliver Gottschalg and Bernd Kreuter¹ discuss possible private equity fund selection criteria and their efficiency.

Given the rather disappointing average historic performance of private equity and the high performance differences across private equity fund managers, accurate fund selection is as crucial for the performance of one's private equity portfolio as it is challenging. Data paucity, limited benchmarking possibilities and the long time lag between commitment decisions and performance outcomes makes private equity fund due diligence still look more like an art than a science. This article assesses the efficiency of commonly used criteria to select fund managers based on historic data. Drawing on a comprehensive analysis of 615 historical due diligence situations, we document the relationship between GP characteristics (measured at the time a new fund is raised) and the subsequent performance of that *focal* fund. We are looking at various measures of past GP performance, but also at other aspects, such as dealflow or experience and assess to what extent these are statistically significant determinants of focal fund performance. In a second step, we assess the *selection efficiency* of different criteria, i.e. the degree to which their use for fund selection purposes would have historically led to above-average

portfolio performance. Our results point to the limited efficiency of 'generic' selection rules, such as the 'top quartile' rule, especially compared to more comprehensive fund rating approaches that simultaneously consider multiple complementary selection criteria.

THE RESEARCH APPROACH

The dataset used for this study contains detailed (anonymous) information on a large sample of North American and European private equity funds²: (1) historical cash inflows and outflows (including fees), (2) historical net asset values of unrealised investments, (3) vintage year, committed capital and geographic focus of the fund and (4) the size (equity value), stage and industry of the underlying investments made by these funds.

From this data 615 historic fundraising situations have been replicated as follows. First, 615 'focal funds' raised in 1999 or before were selected. For these funds, actual performance as of today can already be measured with a sufficient degree of accuracy. For each of the 615 focal funds, data has been composed to reflect the characteristics of the managing GP at the moment of the fundraising, similar to the information that would have been available to a potential investor in the fund at that time.

RELEVANT GP CRITERIA

The 615 simulated due diligence assessments were based on the following

QUANTITATIVE FUND DUE DILIGENCE

information: (a) data on the 'latest mature' fund, i.e. the last fund the focal GP has raised prior to the focal fund which is at least four years old (again to make sure performance information on this fund was reliable at the moment of the hypothetical due diligence), (b) data on the entire track record of the GP, including the past performance of all prior funds of the same GP, (c) GP-level variables, such as GP experience or dealflow and finally (d) data on how the focal fund differs from its most recent predecessor fund. Figure 1 (p. 96) illustrates how data for the hypothetical historic fundraising situations has been composed. Based on this data, a number of distinct measures were constructed.

Performance track record

As the most widely used – and presumably most important – due diligence criterion, we put heavy emphasis on the analysis of the GP's performance track record. It is important to keep in mind that all performance data from prior funds is measured as of the beginning of the vintage year of the focal fund, as this snapshot would have been relevant for focal fund due diligence purposes. The final performance of these funds when they reached their liquidation age may differ from this intermediate performance snapshot. We calculate standard performance measures, such as IRR and Performance Quartiles, as well as the 'Delta IRR', i.e. the difference between actual IRR and the average IRR of a fund's same-vintage and same-stage peers. We considered either the 'latest mature' fund or the average of all prior funds³.

1. Oliver Gottschalg is Assistant Professor at HEC School of Management in Paris, co-director of the HEC-INSEAD Buyout Research Program and Head of Research at Peracs Ltd. Dr. Bernd Kreuter is Head of Private Equity at Feri Institutional Advisors GmbH in Bad Homburg, Germany. The authors can be reached at gottschalg@hec.fr. Background information regarding this research is available at www.buyoutresearch.org.

2. The authors would like to thank Thomson Venture Economics, Gemma Postlethwaite and Jesse Reyes, as well as several research partners from the community of investors into private equity funds who have chosen to remain anonymous, for making this project possible through generous access to their databases.

3. Whenever performance data from several prior funds is used, their performance is aggregated by weighting funds by both their size and their duration. This is the closest possible approximation of the overall performance of the GP.

Dealflow

Another important aspect to look at is the ability of a GP to generate an appropriate and stable flow of investments. This ability can be assessed using two complementary measures. First, the *Percentage of Fund Size Invested* (measured as of year 4 after inception) for the 'latest mature' fund. This variable captures if the GP was able to find enough investments opportunities to invest the capital raised in the most recent mature fund. Second, the *Variance in Number of Deals* per year of the GP prior to focal fund vintage, which measures whether investments occurred regularly or in waves, where the latter could be interpreted as a possible indication of lower dealflow generation ability.

GP experience

Experience is measured through two alternative variables. First the *number of prior funds* raised by the GP and second as the count of the *number of prior investments* made by the GP prior to the focal fund's vintage (including multiple investment rounds).

Differences between the focal and prior funds

The relevance of past performance as an indicator of future fund performance is expected to decrease if focal fund characteristics differ from those of previous funds. Particularly relevant in this context are changes in fund size. We capture this effect by including the *Percentage Change in Fund Size* between focal fund and latest mature predecessor fund in the analysis

WHICH FACTORS CORRELATE WITH FUTURE PERFORMANCE?

A bivariate correlation analysis shown in Table 1 documents which of the different GP characteristics are significantly correlated with the ultimate performance (IRR) of the focal fund. Several observations are in order. First of all, we find support for the view that

measures of past performance of a GP's funds (as of the vintage year of the focal fund) are strongly correlated with the subsequent performance of the next fund raised by this GP. Interestingly, measures of relative performance (Latest Mature Delta IRR, Overall Weighted Delta IRR, Overall Weighted Quartile) show stronger correlations than comparable measures of absolute performance (Latest Mature IRR, Overall Weighted IRR).

This suggests that performance persistence is driven by a GP's ability to repeatedly generate returns that are higher than those of a peer group of comparable funds, rather than to always generate returns of the same magnitude. In other words, even high performing GPs are influenced by exogenous factors that create particularly attractive or difficult investment conditions in a given period and segment of the market. At the same time, the bivariate analysis also shows support for the importance of GP experience as a determinant of future returns of the focal funds: funds raised by GPs with either a larger number of prior funds or a larger number of prior deals perform better *ceteris paribus*.

RANDOM CHOICE VS. THE CRYSTAL BALL: AN APPROACH TO MEASURING PE FUND SELECTION EFFICIENCY

The preceding statistical analysis shows which GP characteristics are

significant determinants (in statistical terms) of focal fund performance. The *economic* relevance of these potential fund selection criteria is a related, yet different question. In other words, we still need to assess the suitability of different GP characteristics to select a portfolio of PE funds that performs better than a random fund choice.

To this end, we first determine the upper and lower 'benchmark performance' values for alternative fund selection rules. We calculate the average return of all 615 focal private equity funds in our sample. This lower bound benchmark corresponds to the return an LP would have enjoyed had she invested proportionally in all private equity funds offered – or in a random sub-set of those. We find a benchmark performance of 17.13 percent simple average IRR or 13.26 percent weighted average IRR for an investor investing in all 615 proposed funds with a total portfolio size of \$212 billion. Any *efficient* fund selection rule should be able to lead to an average performance above this value.

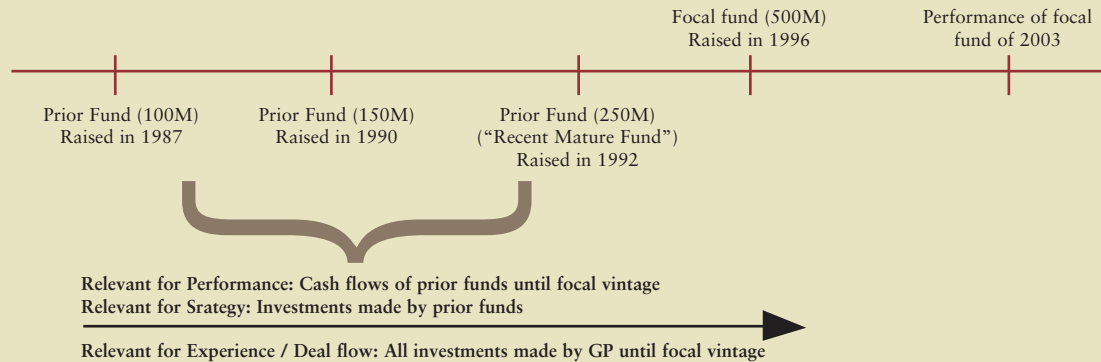
To assess *how efficient* different criteria are, it is also important to assess the distribution of returns in the fund population. In other words we need to know the aggregate performance of the best 10 percent, 11 percent, 12 percent of funds in the population and so forth. We determine these values by ranking all focal funds in the population by their end-

TABLE 1: CORRELATION WITH FOCAL FUND IRR

	Correlation Coefficient
Latest Mature IRR	0.111(**)
Latest Mature delta IRR	0.180(**)
Latest Fund % Inv. Year 4	-0.045
Overall Weighted IRR	-0.008
Overall Weighted Delta IRR	0.103(*)
Overall Weighted Performance Quartile	0.126(**)
Change in Fund Size since Latest Mature Funds	-0.066
Number of Prior Funds	0.137(**)
Number of Prior Deals	0.160(**)
Variance in Deals per Year	-0.020

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

FIGURE 1: EXAMPLE – COMPOSING DATA FOR HYPOTHETICAL FUND RAISING EVENTS

The hypothetical due diligence situations assess the profile of a GP offering the focal fund based on information available at the moment of fundraising and compare these to the actual ex-post performance of the focal fund as of 12/2003.

of-life performance and plot the cumulative average performance of the best 10 percent, 11 percent, 12 percent etc. of these funds relative to the 17.13 percent average as the purple line in Figure 2 (p. 97).

The corresponding line can be interpreted as the result of a 'crystal ball' fund selection device through which an investor would have perfectly foreseen the future performance of each focal fund at its vintage and invested accordingly. This selection device is obviously impossible to realize, as the exact future performance of proposed funds is unknown ex ante. However it constitutes a good upper benchmark in terms of selection efficiency that alternative selection schemes can be compared to. For example, an LP with the 'crystal ball' could have directly selected the best performing 22 percent of funds with an average performance improvement of 56 percent. Similarly the best performing 57 percent of funds had an average performance improvement of 20 percent over the average performance of 17.13 percent.

THE PERACS PRIVATE EQUITY SELECTION EFFICIENCY MEASURE™

We can use these previously developed upper and lower benchmarks to assess and compare the selection effi-

ciency of different fund selection rules in the following way. If we take a given criterion (for example past performance) and apply it to the historic data to select, for example 20 percent of the overall population, we can compare the average performance of this choice to the true top 20 percent of the entire population. Based on this intuition, the 'PERACS Private Equity Selection Efficiency Measure™ (PESEM™) allows us to comprehensively quantify and compare the selection efficiency of different fund selection methods. PESEM™ is defined as the ratio of (a) the integral of the difference between the average performance of all PE funds offered to investors and the average performance of the best x percent of the PE funds as predicted by the selection method over (b) integral of the difference between the average performance of all PE funds offered to investors and the crystal-ball line, i.e. the average performance of the actual best x percent (ex-post) of the PE funds offered to investors. PESEM™ takes values close to 100 percent if the efficiency of the assessed method approaches the performance of the 'crystal ball' portfolio and tends towards 0 for methods that only offer average performance. Should a selection method point to below-average

funds, PESEM™ turns negative.

The PESEM can be interpreted as follows: a PESEM of 50 percent enables investors (on average) to reach a level of performance improvement over the average portfolio equivalent to half the improvement that a true crystal-ball device would have generated. In the following, we illustrate the use of this method based on popular fund selection criteria.

THE SELECTION EFFICIENCY OF PERFORMANCE-BASED FUND SELECTION RULES

The arguably most 'generic' fund selection rule corresponds to the common wisdom of 'backing only top-quartile GPs'. Had an LP selected only funds of GPs whose most recent mature fund rank in the top quartile of their relevant peer group, she would have invested \$99 billion in a portfolio of 216 funds with a weighted average IRR of 16.41 percent. Had she chosen to also include funds with mature predecessor funds in the 2nd performance quartile, she would have invested \$158 billion in a portfolio of 216 funds with a weighted average IRR of 13.66 percent. It is striking that the rule of selecting funds from the 'upper two quartiles' of their respective peer group improves weighted portfolio performance (then 13.66 percent IRR) relative to the benchmark of

random investment (13.26 percent IRR) by only 40 basis points.

A slightly more sophisticated version of a past performance based selection scheme ranks all focal funds by the weighted average IRR of all their predecessor funds and invests into the top percent of funds according to this ranking. We assess the performance of the best 10 percent, 11 percent, 12 percent etc. of funds according to this list and plot the results as the blue line in Figure 2. We note that surprisingly, selection schemes based only on past GP performance were historically not very efficient at identifying a high-performing portfolio. In line with what has been indicated already in the analysis of quartile-rules as selection criteria, we have to conclude that selection schemes that are based on past GP performance only do not make it possible to improve the average portfolio performance much above the lower benchmark of average portfolio performance. It is also interesting to note that this particular selection scheme does not generate a monotonous relationship between the supposedly best x percent selected and the performance of this selection, as can be seen from the peak of the graph in the area of about 30 percent of funds selected. At best, this past performance based selection rule makes it possible to

generate average portfolio returns of 26.4 percent IRR for a portfolio size of 28 percent of the proposed funds. This optimum point for the past performance based selection rule looks like a substantial improvement over the average portfolio performance (17.3 percent simple average IRR), but remains substantially below the 'crystal ball' upper benchmark of over 63 percent average IRR for the same number of funds.

The efficiency of the past-performance-based selection rule can now be illustrated in Figure 2. The PESEM for past-performance-based selection is the ratio between the area below the blue line and the area below the purple crystal ball line in Figure 2, which corresponds to a value of 2 percent. Hence investors using this rule reach a level of performance improvement over the average portfolio that corresponds to 2 percent of the power of a crystal-ball device.

INGREDIENTS OF AN EFFICIENT FUND SELECTION MODEL

The natural next question becomes: is it possible to construct a fund selection model that comes closer to the crystal ball than methods based on past performance only? Our research shows that this is indeed feasible. Key to improving fund selection is the correct combination of multiple criteria. One concrete

example is a proprietary fund selection model that has been jointly developed by the due diligence advisory firm PERACS and the European LP Feri Institutional Advisors. It is based on a multifactor fund rating metric that combines different measures of Performance track record, dealflow, GP experience and differences between the focal and prior funds. We tested this model on the 615 historic fundraising events in our data and the selection rule increased portfolio performance substantially. The yellow line in Figure 2 compares the performance of the portfolio of the best x percent of funds selected by this fund rating model to the performance of the crystal ball upper benchmark, as well as to the previously used past-performance-based selection results.

The chart illustrates the efficiency of the fund rating model along the entire range of selected portfolio sizes. Historically, this method would have enabled an investor to select a \$73 billion portfolio of funds (1/3 of the population) with twice the average performance or the best 20 percent of funds with an average performance of over 45 percent average IRR. Even for the 28 percent of selected funds for which the past-performance-based method offered the best results, the fund rating model leads to much better results (21 percent vs. 9 percent average IRR improvement of the selected funds).

This multi-factor fund rating model has a PESEM of 35 percent, in other words it enables investors to reach a level of performance improvement over the average portfolio equivalent to 35 percent of the improvement that a crystal-ball choice would have generated. While a true crystal ball remains impossible to construct, this approach shows that it is both possible and worthwhile to make some progress towards building something similar. ■

FIGURE 2: SELECTION EFFICIENCY OF FUND SELECTION SCHEMES

