PitchBook

PitchBook Data, Inc.

John Gabbert Founder, CEO

Nizar Tarhuni Vice President, Institutional Research and Editorial

Daniel Cook, CFA Head of Quantitative Research

Institutional Research Group

Analysis

Andrew Akers, CFA
Senior Quantitative Research Analyst
andrew.akers@pitchbook.com

Andy White
Lead Analyst, Quantitative Research
andy.white@pitchbook.com

Zane Carmean, CFA, CAIA
Lead Analyst, Quantitative and Funds Research
zane.carmean@pitchbook.com

pbinstitutionalresearch@pitchbook.com

Publishing

Designed by **Drew Sanders**

Published on July 31, 2023

Contents

PitchBook Manager Performance Scoring

PitchBook is a Morningstar company providing the most comprehensive, most accurate, and hard-to-find data for professionals doing business in the private markets.

Introduction

The PitchBook Manager Performance Scores are a quantitative framework designed to assess the performance track record of a fund manager's closed-end private market strategies, also known as fund families. Manager Scoring is designed to allow for more accurate comparisons between general partners (GPs) despite individual funds being launched at different times and investing in different economic environments. This is primarily accomplished by converting each fund's internal rate of return (IRR) to a Z-score, controlling for the strategy and vintage of the fund. Individual fund Z-scores are then weighted by the estimated finality of the IRR and combined across the fund family to produce a single score. This document provides the methodological details of the PitchBook Manager Performance Scores, including benchmark selection, performance measurement, fund weighting, and the ultimate scoring of fund families.

Benchmark selection

Each fund in the family is benchmarked against its peers based on the median performance of its peer group. For peer group selection, we segment the fund universe on two important factors to fund performance: vintage year and fund class. Vintage year cohorts are constructed from three-year rolling windows in which the fund in question has a vintage year in the middle of the window. For example, the vintage cohort for a 2010 vintage fund is 2009 to 2011. This is done to account for the fact that funds of proximate vintages are investing in largely similar economic environments, thus providing a logical benchmark for one another. Funds are also segmented into seven broad classes: private equity, venture capital, real estate, real assets, debt, secondaries, and funds of funds.

- **Key consideration:** We use rolling three-year vintages for our benchmarks selection to broaden the data set that a fund is compared with. Some vintages have a wider dispersion of IRRs than others, with varying fund counts; the rolling benchmark helps smooth out the dispersion of relative performance.

Performance scoring

Performance is measured via each fund's IRR, which is first normalized by subtracting the fund's IRR from the median IRR of its benchmark to get an excess return. Next, we calculate a modified Z-score of the excess return for fund *i* of class *j*, where $\widetilde{EIRR}$ is the median excess IRR and *MAD* is the median absolute deviation of excess IRRs:

$$z_{ij} = \frac{0.6745\left(EIRR_i - \widetilde{EIRR}_j\right)}{MAD_j}$$

We use the modified Z-score because it is more robust with respect to outliers than the standard Z-score that uses mean and standard deviation.[1] It is important to note that median excess IRR and median absolute deviation of excess IRRs are based on funds of the same class, regardless of vintage year. Since we have already normalized the IRR relative to the vintage benchmark, we are free to compare funds of the same class across all vintages. Thus, the Z-score for fund *i* can be interpreted as how good or bad its excess IRR was compared with its respective fund class. The Z-scores are then winsorized at values more extreme than ±3.5 to mitigate the impact any single fund can have on a strategy's track record.

- **Key consideration:** We use the modified Z-score because it is more robust with respect to outliers than the standard Z-score that uses mean and standard deviation. It has the added benefit of interpretability with a negative score representing the bottom 50% of fund IRRs and a positive score representing the top 50% of fund IRRs. Because each Z-score is a measure of performance relative to the benchmark peer group, we are free to compare funds across vintages.

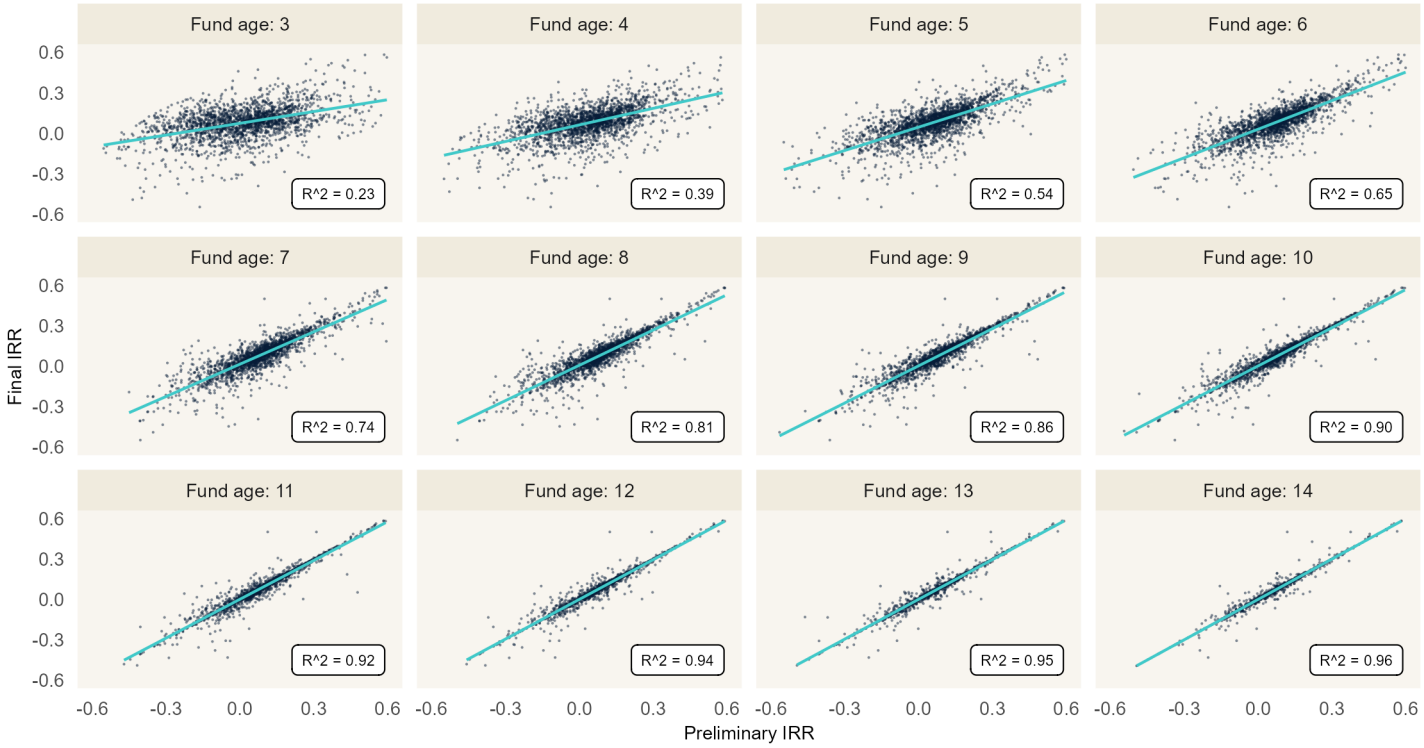
Traditionally, closed-end fund benchmarking is done on a fund-by-fund basis. It is important to control for the economic environment in which the fund is operating; thus, fund vintage year provides important context. However, there is no standard way to measure a fund manager's performance over the course of a series of funds, which have invested in different environments. Commonly, LPs use fund IRR quartiles across a series of funds to measure performance persistence, but this is a blunt measure because it does not account for the degree of outperformance or underperformance of an individual fund. For example, for 2003 vintage PE funds, the difference between median IRR and top-quartile IRR is 12%. For vintage 2006 funds, that difference is only 4%. Thus, even funds in the same quartile can have very large or very small absolute variance from one another. PitchBook Manager Scoring accounts for these nuances by leveraging Z-scores rather than a simple quartile metric.

1: *How to Detect and Handle Outliers*, Boris Iglewicz and David C. Hoaglin, American Society for Quality Control, 1993.

Fund weighting

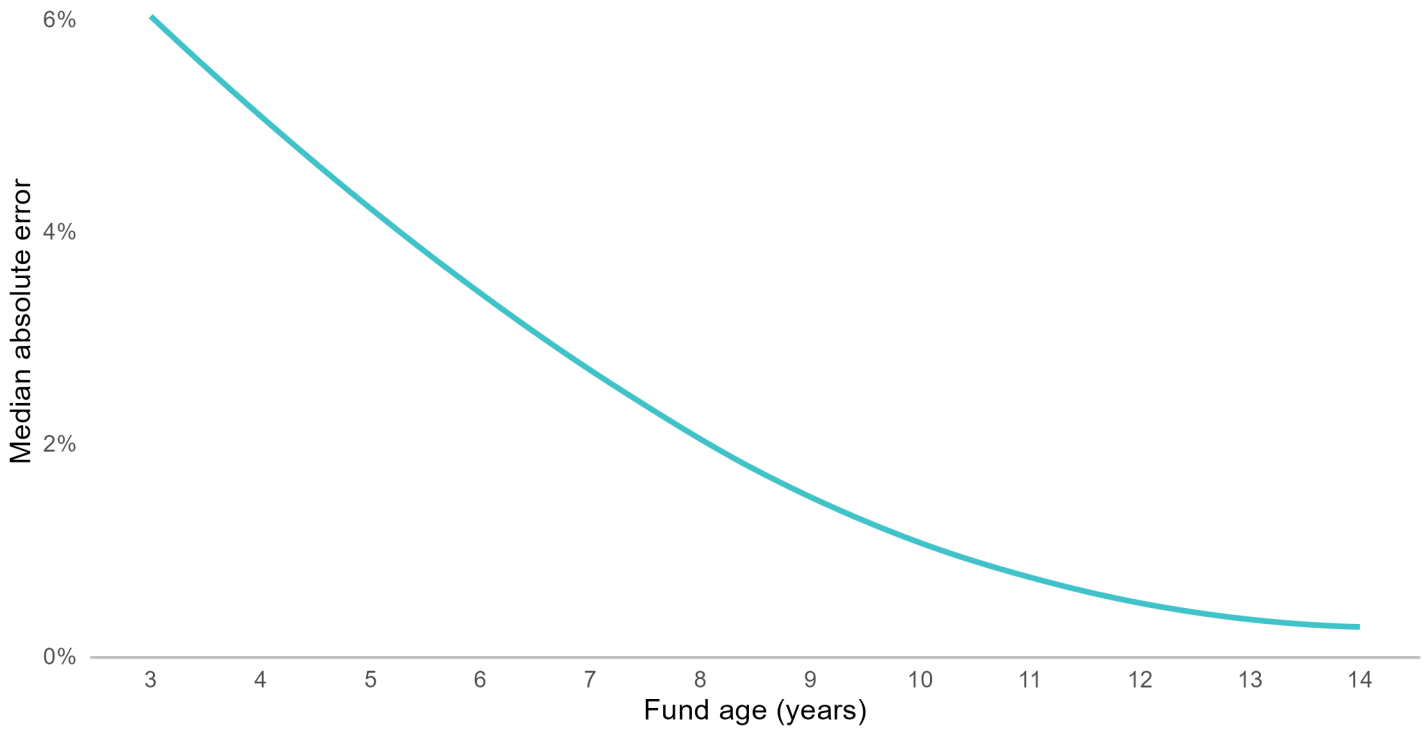
Closed-end fund performance is only known with certainty after all investments have been liquidated. Interim IRR reports are particularly prone to being misleading, and very early reports are often not indicative of the final returns figure. As such, to create a single score for a fund family we need a weighting scheme that takes time-based uncertainty into account. For starters, we must determine at what age funds should be included in the track record and at what age a fund's performance should be considered "final." A simple method to guide this decision is to look at the historical data to determine how predictive interim performance is of final performance as a function of fund age. The first chart below shows the relationship between interim IRR and final IRR for fund ages of 3 to 14 years. The second chart shows the median absolute difference between the fitted line and the final IRR at each fund age.

Relationship between interim and final IRR figures by fund age*



Source: PitchBook • Geography: Global
*As of December 31, 2022

Median absolute difference between interim and final IRR figures by fund age*



Source: PitchBook • Geography: Global
*As of December 31, 2022

As expected, the goodness of fit improves as funds get older. At age 4, for example, a fund's preliminary IRR is only mildly predictive with an R^2 of 0.40 and a median absolute error of 5.2%. By age 5, these measures improve significantly to 0.54 and 4.3%, respectively. Although the decision is somewhat arbitrary, we believe it is reasonable to include a fund in a strategy's performance track record once it reaches the age of 5. On the other hand, we need to determine a general ex-ante cutoff for when fund performance is considered final. For this, we looked at both the absolute predictive power as well as the change in predictive power of preliminary IRR as fund age increases. By age 12, the change in median absolute error begins to slow as it falls further below 1.0%. For example, the net decline in median absolute error from age 8 to age 10 is 1.0%, while the net decline from age 10 to age 12 is just 0.4%. Based on these observations, we consider the performance of funds 12 years or older as final.

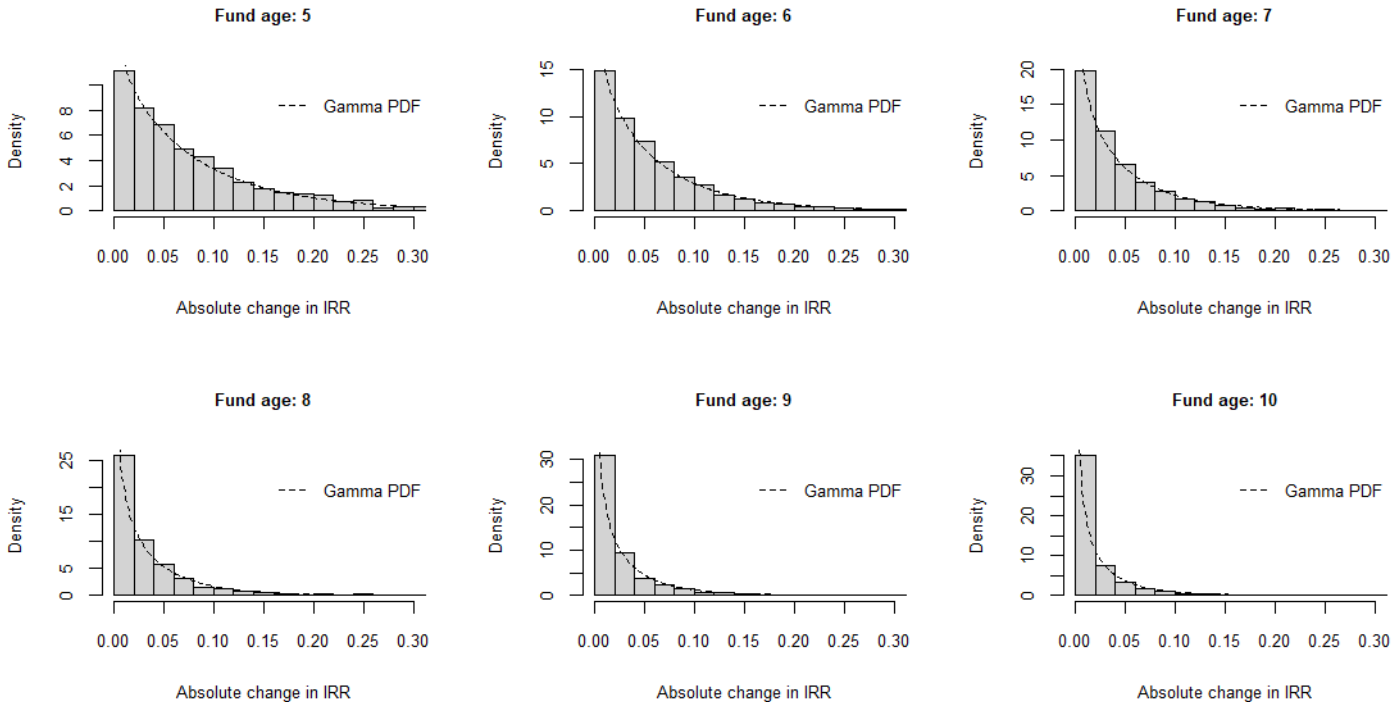
- **Key consideration:** Fund IRRs change as investments are made, cash is generated, and investments are sold through the fund life. Early on, interim IRRs are not a good indication of where the fund's eventual IRR will end up. As the fund matures, its IRR becomes more "settled."

Next, we need to determine the weights of each fund included in the family. Generally, funds closer to having final performance numbers (that is, closer to 12 years old) should be given higher weights in the track record calculation for a strategy. A good starting point would be to estimate uncertainty solely based on fund age. As demonstrated above, age is an important factor in determining how much performance is expected to change from the current value. However, age alone does not account for the significant variation observed across individual funds; thus, additional data should be incorporated. In particular, the cash flow characteristics of a fund are key indicators of the uncertainty in its performance figures. Distributions, for example, represent a portion of the IRR calculation that has already been solidified. All else being equal, a fund with a higher ratio of distributions to paid-in (DPI) capital should have less uncertainty.

Modeling performance uncertainty—defined as the expected absolute change in IRR—requires examination of its distributional characteristics. Two main characteristics lead us to leveraging a generalized linear model (GLM) rather than a linear regression model. First, absolute change in IRR must be greater than zero. Second, it is heavily positively skewed, meaning that most values are relatively small and larger values are increasingly unlikely. Both these characteristics make linear regression unfit for our purposes. This is the case for two reasons: 1) the assumption that the regression residuals are normally distributed is violated, and 2) the predictions from a linear regression model can be negative.

Although a GLM is akin to linear regression in that it fits a coefficient for each covariate, rather than assuming the residuals are normally distributed, it makes a distributional assumption about the response variable. The GLM then fits the model parameters by maximizing the likelihood of the observations based on the probability density function (PDF) of the assumed distribution. Therefore, we first need to determine if the absolute change in IRR can be modeled using a known probability distribution. Based on maximum likelihood estimates, the best fit is the gamma distribution, which matches the skew of the data and is also constrained to positive values. The plot below displays the empirical distribution versus the fitted gamma distribution for several fund ages.

Empirical distribution and fitted gamma PDF of absolute change in IRR by fund age*



Source: PitchBook • Geography: Global
*As of December 31, 2022

With the distribution selected, the model can be expressed as follows:

$$\max_{\theta} \sum_{i=1}^{n} L\left(\beta e^{\theta^T x_i}, \beta\right)$$

L is the log likelihood function of the gamma distribution and $\beta e^{\theta^T x_i}$ and β are the rate and shape parameters, respectively. The term $\theta^T x_i$ is the inner product between a vector of fitted coefficients and the features of a single data point, which is the same interpretation as in traditional linear regression. Given that the mean of the gamma distribution is the rate parameter divided by the shape parameter, we can see that the mean, and expected value, for observation *i* is $e^{\theta^T x_i}$.

The GLM allows us to directly model the relationships between individual fund characteristics other than age and the uncertainty of preliminary performance, which is particularly important for younger funds because there is more variation across funds. We find that there are two important characteristics that impact uncertainty: 1) a relative measure of distributions, such as DPI, and 2) the absolute deviation of the IRR from the mean. DPI is relevant to uncertainty because it represents a proxy of the percentage of cash flows that have already been realized in the IRR calculation, and higher DPI values lower the probability of significant future distributions. The second factor, IRR deviation from the mean, captures the widely recognized idea of mean reversion: That is, funds with high preliminary IRRs typically see them decline as time goes by and funds with low preliminary IRRs typically see them increase.

Armed with these insights, we fit a separate GLM for each fund age in quarterly increments. For all fund ages, we find that DPI and IRR deviation are both statistically significant predictors of uncertainty, although the relationships are weaker as funds get older. As an example, the table below shows the model summary for funds that are 8 years old. Uncertainty, $\hat{y}$, can then be calculated as: $\hat{y} = exp(-3.26 + 3.22Dev - 0.76DPI)$.

GLM summary for 8-year-old funds

	Beta	t-statistic	p-value
Intercept	-3.26	-58.63	0.00
IRR deviation	3.22	11.36	0.00
DPI	-0.76	-13.92	0.00

- **Key consideration:** We prefer to weight more certain returns higher than less certain returns; thus, the absolute weight is simply the inverse of predicted uncertainty. Because IRR is taken as an absolute value, the weight assigned to a fund in the family score increases as the IRR approaches the mean, as DPI increases, and as the fund ages.

The last step in the weighting process is to assign a prior assumption that a fund's performance becomes final when it reaches 12 years old. The strength of this assumption increases linearly as funds age. This means that a fund's weight is pulled toward the maximum weight as each quarter passes and ensures that all funds older than 12 have equal weighting. This step has two additional benefits: 1) it respects the dominant relationship between fund age and performance uncertainty, and 2) it provides smooth fund weightings from quarter to quarter such that there are not large changes in weightings when a fund reaches age 12. The weight of the prior increases linearly each quarter from 0% at age 5 to 100% at age 12. Because there are 28 quarters in this timespan, the weight of the prior increases by approximately 3.6% each quarter.

Putting all the pieces together, the absolute weight of fund *i* of age *n* is calculated as:

$$W_i = max\left(\bar{u}_p,\ w_n\bar{u}_p + (1 - w_n)e^{\theta_n^T x_i}\right)^{-1}$$

Where:

$\bar{u}_p$ = the prior uncertainty (the average uncertainty for funds aged 12)
w_n = the weight of the prior at fund age *n*
θ_n = the vector of GLM coefficients for fund age *n*
x_i = the vector of data for fund *i*

The relative weight of a fund *i* within strategy *j* is then just its proportion of the sum of all weights:

$$w_i = \frac{W_i}{\sum_{j=1}^{n} W_j}$$

With weights and scores for each fund, the final track record score for strategy *j* is a weighted average over all eligible funds in the strategy:

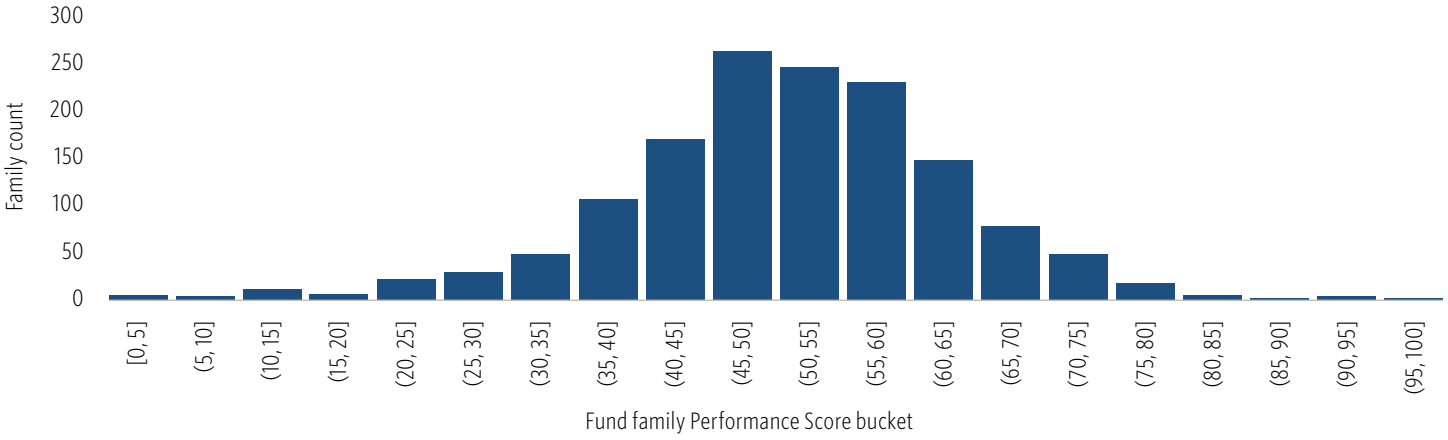
$$Z_j = \sum_{i=1}^{n} w_i \, z_i$$

Finally, the scores are normalized on a 0-100 scale. The +3.5 and -3.5 terms reflect the fact that we have winsorized the fund Z-scores at those levels earlier in the process, meaning they represent the maximum and minimum possible family Z-score.

$$Performance\ Score_j = \frac{Z_j + 3.5}{3.5 - (-3.5)} \times 100$$

- **Key consideration:** In summary, the weighted average score for a fund family takes into account the uncertainty of each constituent fund's interim IRRs. The older the fund (up to 12 years), and the more DPI the fund has achieved, the higher likelihood that the interim IRR is predictive of the final IRR, thus the greater weight assigned in the overall score calculation. The 0-100 scale means that 50 represents a neutral performance score (neither beating nor losing to the fund family's respective benchmarks).

Dispersion of fund family Performance Scores across all strategies*



Source: PitchBook • Geography: Global
*As of December 31, 2022

Eligibility

In addition to individual fund eligibility criteria, strategies are required to have at least two eligible funds to be eligible for a track record calculation.

Firms must also be active to receive a final score. Active is defined as having raised a fund within the prior 10 years, and the firm's current status indicates that it has not ceased operations (as in, has not gone out of business). However, data for firms that do not receive a final score is used in the broader calculation and included in the benchmarking process to reduce survivorship bias.