

Survivorship Bias in the Cross-Section of Winner Portfolios

Ayush Jha¹ Ali Jaffri² Hassan Butt³

¹Texas Tech University

²North Dakota State University

³Missouri Southern State University

Abstract

Backtests of past winners often use databases that retain firms that survive, but it is unclear whether the resulting bias reflects changes in which firms are selected or the omission of outcomes for firms that later exit. We separate these channels using point-in-time data on U.S. common stocks and annual winner portfolios formed from 1991 to 2020, with outcomes observed for up to five years. Conditioning the ranking on five-year survival changes one-third of the top twenty names, roughly the amount implied by overall attrition. The resulting survivor-conditioned portfolio nevertheless earns 6 percentage points more over one year and 37 percentage points more over five years than the corresponding full-universe portfolio. An exact decomposition attributes nearly all of this difference to the returns omitted when winning firms exit, rather than to the firms promoted into their place. The effect is concentrated among equal-weighted winners, where distress-related exits are associated with large losses.

JEL classification: G11, G12, G14

Keywords: survivorship bias, performance evaluation, momentum, delisting, point-in-time data.

1 Introduction

Past-performance rankings are often evaluated using records that are incomplete because some of the firms that once ranked highly subsequently disappear. In performance studies, this is the familiar survivorship problem: conditioning on firms that remain in the sample can remove poor subsequent outcomes and raise measured performance. Most evidence on this issue concerns average performance, particularly for mutual funds (Brown, Goetzmann, Ibbotson, and Ross 1992; Elton, Gruber, and Blake 1996; Carhart, Carpenter, Lynch, and Musto 2002).

The issue is somewhat different for strategies that select firms from the upper tail of past returns. Here, conditioning on survival can affect both the composition of the winner set and the subsequent returns observed for firms that remain in it. We call these channels *composition* and *truncation*. Composition arises when firms that exit are removed before the ranking is formed, allowing other firms to enter the winner set. Truncation arises when the database omits the subsequent returns of winners that exit. Separating the two matters because they describe different sources of bias: one concerns which firms are selected, while the other concerns which outcomes are observed.

We study these channels using point-in-time U.S. equity data. At each annual formation date from 1991 through 2025, we rank common stocks by their trailing twelve-month returns using either the full set of firms active at formation or only firms that remain listed at a specified future horizon. The latter ranking is a research counterfactual: it represents the ranking that would be obtained if a backtest were conducted using a survivor-only database. We then compare portfolios formed from the two rankings and construct portfolios of firms common to both rankings and firms promoted only after survival conditioning.

The comparison has a simple accounting structure. Because both rankings use the same past return measure, a full-ranking winner that survives remains among the top K firms after surviving firms are re-ranked. Each full-ranking winner that exits therefore corresponds to one promoted firm. The fraction of winner slots that changes hands is consequently equal to the attrition rate among the original winners. This identity provides a direct benchmark for interpreting reclassification: a high rate of replacement need not indicate that survival is informative about which firms were better past winners.

The portfolio comparison yields a different result. At the five-year horizon, conditioning on survival changes roughly one-third of the top twenty names, close to the attrition rate for the full universe. Yet the equal-weighted survivor-conditioned portfolio earns 6.2 percentage points more over twelve months and 37.4 percentage points more over sixty months than the portfolio formed from the full ranking. The exact decomposition shows that most of this difference comes from the returns of winners that subsequently exit. The promoted firms do not, on average, perform sufficiently differently from the common winners to account for the portfolio gap.

The exit data provide a reason for this result. Total five-year attrition among the top twenty winners is close to the market-wide rate, but the types of exits differ. Distress delistings are more common among

the top twenty than in the full universe, while acquisitions are less common. Because distress exits are associated with large losses before delisting, removing these observations can materially raise the measured performance of a winner portfolio even when the number of exits is not unusual.

The evidence is specific to the setting studied. The effect is concentrated in equal-weighted portfolios and is smaller under value weighting. It is also larger in the earlier part of the sample. With only 30 formation dates for the five-year analysis, estimates of time variation are necessarily imprecise. We therefore emphasize the accounting decomposition and date-level and permutation-based inference rather than interpreting individual cross-sectional interaction estimates as evidence of broad state dependence.

1.1 Related Literature and Contribution

The paper relates to several strands of work on survivorship, data construction, and return measurement.

Survivorship and selection in performance evaluation. Survivorship bias has been studied extensively as a source of bias in measured average performance, especially for mutual funds (Brown, Goetzmann, Ibbotson, and Ross 1992; Elton, Gruber, and Blake 1996; Carhart 1997; Carhart, Carpenter, Lynch, and Musto 2002). Related work considers attrition in performance-persistence tests (Carpenter and Lynch 1999; Linnainmaa 2013), reporting incentives in hedge-fund databases (Aiken, Clifford, and Ellis 2013), and selection in private-market return data (Cochrane 2005; Korteweg and Sorensen 2010). Our focus is narrower: strategies that select firms using past returns. We show how the resulting survivorship difference can be separated into a change in the selected firms and the omission of subsequent outcomes for firms that leave the sample.

Data construction and cross-sectional return evidence. A related literature studies how choices in data construction affect cross-sectional return evidence, including multiple testing, replication, weighting, and the treatment of delisted securities (Harvey, Liu, and Zhu 2016; Harvey and Liu 2020; Hou, Xue, and Zhang 2020; Chen and Zimmermann 2022; Jensen, Kelly, and Pedersen 2023; Shumway 1997; Shumway and Warther 1999; Beaver, McNichols, and Price 2007). Recent work also examines revisions to the CRSP data and their implications for asset-pricing results (Schwarz, Walter, and Weiss 2026). Our analysis holds the formation rule fixed and changes only whether the universe is conditioned on later survival. The comparison measures the performance difference associated with that data choice.

Winners, distress, and subsequent returns. The upper tail of past returns is central to momentum and related winner-selection strategies (Jegadeesh and Titman 1993; Hendricks, Patel, and Zeckhauser 1993; Grinblatt, Titman, and Wermers 1995). Previous studies document links between momentum, firm distress, and the distribution of stock returns (Avramov, Chordia, Jostova, and Philipov 2007; Campbell, Hilscher, and Szilagyi 2008; Daniel and Moskowitz 2016; Conrad, Kapadia, and Xing 2014). Our exit

analysis complements this work by examining how the type of subsequent exit varies across the trailing-return distribution. In the upper tail, distress exits are more frequent than in the market as a whole, while acquisitions are less frequent. This difference in exit type helps explain why a survivor-only record can produce a higher winner-portfolio return without a correspondingly higher rate of attrition.

Expectations and performance information. Past returns also enter models of return expectations and settings in which investors evaluate managers or firms using recent performance (Greenwood and Shleifer 2014; Barberis, Greenwood, Jin, and Shleifer 2015, 2018; Bordalo, Gennaioli, La Porta, and Shleifer 2019, 2024; Kaniel and Parham 2017; Evans and Sun 2021; Ben-David, Li, Rossi, and Song 2022). This motivates a secondary question about whether the survivorship measure varies with market states associated with extrapolative expectations. The available formation-date variation provides little evidence of a stable state-dependent relation.

The rest of the paper is organized as follows. Section 2 sets out the accounting framework. Section 3 describes the data, and Section 4 states the estimating equations and inference procedures. Section 5 measures reclassification, Section 6 decomposes the portfolio difference, and Section 7 examines exit rates across the return distribution. Section 8 reports robustness checks, and Section 9 concludes. The Appendix contains the proof of Proposition 1 and supplementary results.

2 Composition and Truncation

Fix a formation date t and a survival horizon h . Let U_t denote the firms active at t , let R_i denote firm i 's trailing twelve-month return, and let $S_t \subseteq U_t$ denote the firms still listed at $t + h$. The full winner set W_t^F contains the top K firms in U_t ranked by R_i ; the survivor-conditioned winner set W_t^S contains the top K firms in S_t ranked by the same variable. Define

$$\begin{aligned} C_t &= W_t^F \cap S_t && \text{(common winners: winners under both rankings),} \\ E_t &= W_t^F \setminus S_t && \text{(exiting winners: full-ranking winners that do not survive),} \\ P_t &= W_t^S \setminus W_t^F && \text{(promoted firms: winners only after conditioning).} \end{aligned}$$

Proposition 1. *Suppose $|S_t| \geq K$. Then $W_t^S = C_t \cup P_t$ and $|P_t| = |E_t|$. Consequently the reclassification rate $\rho_t \equiv |P_t|/K$ equals the attrition rate among full-ranking winners, $|E_t|/K$. If survival were assigned at random with the observed number of survivors, $E[\rho_t]$ would equal the attrition rate of the universe.*

The result follows from a ranking property: removing firms from a ranking cannot lower the rank of a firm that remains. Thus, every full-ranking winner that survives remains among the top K survivors, while each full-ranking winner that exits creates one vacancy that is filled by a promoted firm. The proposition gives a simple benchmark for interpreting reclassification. A substantial change in the

winner list can reflect substantial attrition, even when the identities of the exiting firms contain no more information about future survival than those of firms elsewhere in the universe.

The same partition gives an exact decomposition of the return difference between the two winner portfolios. Writing R^X for the cumulative return of an equal-weighted portfolio of group X held from t ,

$$\underbrace{R^S - R^F}_{\text{total gap}} = \underbrace{(R^S - R^C)}_{\text{composition}} + \underbrace{(R^C - R^F)}_{\text{truncation}}. \quad (1)$$

The composition term measures the difference associated with replacing exiting winners by promoted firms. The truncation term measures the difference associated with omitting the subsequent returns of the exiting winners. With monthly rebalancing across listed constituents, the corresponding identity also holds month by month: $r_\tau^S - r_\tau^F = a_\tau^P(r_\tau^P - r_\tau^C) - a_\tau^E(r_\tau^E - r_\tau^C)$, where a_τ^P and a_τ^E are the portfolio weights on promoted and exiting firms in event month τ . The weight on exiting firms falls as those firms leave the sample, whereas the promoted firms remain part of the survivor-conditioned portfolio.

This decomposition separates two questions that are often combined in discussions of survivorship bias. Composition concerns whether the firms retained by the survivor-only ranking differ from the firms that would have been selected using information available at formation. Truncation concerns the outcomes that disappear when firms selected at formation subsequently leave the database. The empirical analysis asks how much of the observed portfolio difference is associated with each channel.

3 Data and Design

Sources and universe. We use the WRDS CRSP daily security files in the CIZ format, Compustat North America annual fundamentals linked through the CRSP/Compustat Merged table, and Fama-French factor returns. At each annual formation date in June from 1991 through 2025, the universe includes U.S. common stocks that satisfy the standard share-code and exchange screens and are trading on the formation date. The universe is constructed forward from each formation date rather than from the set of firms present at the time of extraction. Accounting variables are matched using reporting lags so that they are available by the formation date. The resulting panel contains 165,824 firm-date observations on 17,725 securities, or 4,738 firms per formation date on average.

Rankings and exits. Firms are ranked by their trailing twelve-month return. The survivor-conditioned ranking re-ranks firms that remain listed after one year, three years, five years, or through the end of the available sample. The five-year horizon is the primary specification and leaves 30 formation dates from 1991 through 2020 with a complete five-year follow-up period. Exits are classified from CRSP delisting codes as distress delistings, acquisitions or mergers, voluntary delistings, and a residual category. Over five years, 33.9 percent of firms exit: 22.0 percent through acquisition, 10.1 percent through distress, and

1.8 percent voluntarily.

Winners and portfolios. Winners are the top K firms by trailing return, with $K \in \{1, 5, 10, 20\}$ and $K = 20$ as the primary definition. We use a fixed number of winners because conditioning on future survival reduces the size of the underlying universe, which would otherwise mechanically change the number of firms selected by a top-decile rule. At each formation date we construct equal-weighted portfolios of the full winners W^F , survivor-conditioned winners W^S , common winners C , and promoted firms P , and hold them for up to 60 months. Portfolio weights are rebalanced monthly among constituents that remain listed. A security contributes its CRSP return through its final month of listing and then leaves the portfolio; no return is assigned after delisting.

Inference. Portfolio and exit-rate estimates are first calculated within each formation date and then averaged across dates. Date-level t -statistics therefore use one observation per formation date. Factor regressions use a non-overlapping calendar-time monthly series with Newey-West standard errors. Firm-level regressions include formation-date fixed effects and use standard errors clustered by security and formation date. Because the five-year analysis contains only 30 formation dates, the main analysis supplements conventional standard errors with permutation and date-level inference. Results involving time variation are interpreted with particular caution.

4 Econometric Specifications

This section collects every estimating equation in the paper; each table note refers back to the equations used. Throughout, $t = 1, \dots, T$ indexes annual (June) formation dates, i indexes securities, $\tau = 1, \dots, 60$ indexes months after formation, and h is the survival horizon ($h = 60$ months unless stated otherwise). $R_{i,t}$ is the trailing twelve-month return at formation, and $L_{t,\tau}$ is the set of securities still listed in month $t + \tau$.

4.1 Rankings, reclassification, and the permutation benchmark

Let $\text{Pct}_{i,t}^F$ be the percentile rank of $R_{i,t}$ within U_t and $\text{Pct}_{i,t}^S$ its percentile rank within the survivor set S_t (ranks in $[0, 1]$, ties averaged). The winner sets W_t^F and W_t^S and the groups C_t , E_t , and P_t are defined in Section 2. The reclassification rate reported in Table 1 is the date average

$$\bar{\rho}_K = \frac{1}{T} \sum_{t=1}^T \frac{|P_t|}{K}, \quad (2)$$

which by Proposition 1 equals $T^{-1} \sum_t |E_t|/K$, the attrition rate among full-ranking top- K firms. The random-attrition benchmark holds fixed, within each stratum c (formation date, or formation date $\times$

GICS sector), the number of firms that exit, $n_{c,t}^{\text{exit}} = |U_{c,t} \setminus S_{c,t}|$. In replication $b = 1, \dots, B$ it draws that many firms uniformly without replacement from $U_{c,t}$, treats them as the exits, re-ranks the remaining firms on $R_{i,t}$, and recomputes (2) as $\bar{\rho}_K^{(b)}$. The one-sided p -value is

$$p = \frac{1}{B} \sum_{b=1}^B \mathbf{1}\{\bar{\rho}_K^{(b)} \geq \bar{\rho}_K\}, \quad B = 1,000. \quad (3)$$

4.2 Portfolio returns, the decomposition, and date-level inference

For a group $X \in \{W^F, W^S, C, P\}$ formed at t , the portfolio return in event month τ is

$$r_{t,\tau}^X = \sum_{i \in X_t \cap L_{t,\tau}} \omega_{i,t,\tau}^X r_{i,t+\tau}, \quad \omega_{i,t,\tau}^X = \begin{cases} 1/n_{t,\tau}^X & \text{(equal weights),} \\ \frac{ME_{i,t+\tau-1}}{\sum_{j \in X_t \cap L_{t,\tau}} ME_{j,t+\tau-1}} & \text{(value weights),} \end{cases} \quad (4)$$

where $n_{t,\tau}^X = |X_t \cap L_{t,\tau}|$ and ME is market capitalization. A security contributes its CRSP return through its final month of listing and then leaves the portfolio; nothing is imputed after delisting. The cumulative return to horizon H is $R_t^X(H) = \prod_{\tau=1}^H (1 + r_{t,\tau}^X) - 1$. For a contrast between groups A and B , let $D_t^{A-B}(H) = R_t^A(H) - R_t^B(H)$. The decomposition (1) holds date by date,

$$\underbrace{D_t^{W^S-W^F}(H)}_{\text{total}} = \underbrace{D_t^{W^S-C}(H)}_{\text{composition}} + \underbrace{D_t^{C-W^F}(H)}_{\text{truncation}}, \quad (5)$$

so the estimated components add exactly to the estimated total. Every contrast is estimated by its date average and tested with one observation per formation date:

$$\bar{D}(H) = \frac{1}{T} \sum_{t=1}^T D_t(H), \quad t(\bar{D}(H)) = \frac{\bar{D}(H)}{s_D(H)/\sqrt{T}}, \quad s_D^2(H) = \frac{1}{T-1} \sum_{t=1}^T (D_t(H) - \bar{D}(H))^2. \quad (6)$$

For $H > 12$ the holding periods of adjacent formation dates overlap, so the $D_t(H)$ are not independent across t . The robustness analysis therefore reports the same statistic on every third and every fifth formation date, for which holding periods of up to 36 and 60 months, respectively, do not overlap.

4.3 Factor-adjusted returns

Because formation is annual and the first twelve months after each June formation date run from July through the following June, the first-year returns of successive cohorts tile the calendar without overlap. Stacking them gives one calendar-time series of $M = 12T = 360$ months for each portfolio. For a contrast $A - B$ we estimate

$$r_m^A - r_m^B = \alpha + \beta' \mathbf{f}_m + \varepsilon_m, \quad m = 1, \dots, M, \quad (7)$$

where $\mathbf{f}_m$ is MKT_m (CAPM); $(\text{MKT}_m, \text{SMB}_m, \text{HML}_m)$ (FF3; Fama and French 1993); these plus RMW_m and CMA_m (FF5; Fama and French 2015); or FF5 plus UMD_m (Carhart 1997). For individual portfolios (Appendix Table A2) the dependent variable is the excess return $r_m^X - r_m^f$. Standard errors are Newey and West (1987) with a Bartlett kernel and four lags, and the reported alpha is annualized as $(1 + \hat{\alpha})^{12} - 1$.

4.4 Exiting winners

Under equal weights the monthly return of the exiting winners still listed is recovered exactly from the full and common portfolios,

$$r_{t,\tau}^E = \frac{n_{t,\tau}^F r_{t,\tau}^F - n_{t,\tau}^C r_{t,\tau}^C}{n_{t,\tau}^F - n_{t,\tau}^C}, \quad n_{t,\tau}^F > n_{t,\tau}^C. \quad (8)$$

Let $\mathcal{M}_t$ be the event months with $n_{t,\tau}^F > n_{t,\tau}^C$. The test statistic in Table 3, Panel B, is the date-level mean

$$d_t = \frac{1}{|\mathcal{M}_t|} \sum_{\tau \in \mathcal{M}_t} (r_{t,\tau}^E - r_{t,\tau}^C), \quad (9)$$

averaged over dates and tested as in (6). The levels of r^E and r^C reported beside it are pooled over all date-months in $\bigcup_t \mathcal{M}_t$.

4.5 Exit rates across the trailing-return distribution

For exit type $k \in \{\text{any}, \text{distress}, \text{acquisition}, \text{voluntary}\}$ and a group g of firms defined by trailing return at t (a decile, the top percentile, or the top twenty), the group exit rate and its deviation from the universe rate on the same date are

$$e_{g,t}^k = \frac{1}{N_{g,t}} \sum_{i \in g_t} \mathbf{1}\{\text{Exit}_{i,t,h}^k\}, \quad \delta_{g,t}^k = e_{g,t}^k - e_{U,t}^k, \quad (10)$$

where $\text{Exit}_{i,t,h}^k$ indicates that i delists for reason k (classified from CRSP delisting codes) within h months. Table 4 reports $T^{-1} \sum_t e_{g,t}^k$ and tests $\delta_{g,t}^k$ with the paired date-level t -statistic (6). The distress share of exits is $T^{-1} \sum_t e_{g,t}^{\text{distress}} / e_{g,t}^{\text{any}}$.

The conditional relation between trailing returns and exit is estimated with the linear probability model

$$\mathbf{1}\{\text{Exit}_{i,t,h}^k\} = \gamma_t + \beta_1 \text{Pct}_{i,t}^F + \beta_2 R_{i,t} + \delta' \mathbf{X}_{i,t} + u_{i,t}, \quad (11)$$

with $\text{Pct}_{i,t}^F$ in percentile points (0–100), formation-date fixed effects γ_t , and controls $\mathbf{X}_{i,t}$: log market capitalization, log firm age, log book-to-market, 36-month market beta, idiosyncratic volatility, Amihud (2002) illiquidity, leverage, and profitability, all measured at formation with accounting data lagged to

be publicly available. The model is estimated separately for $h \in \{12, 36, 60\}$ months on the formation dates for which h is fully observed.

4.6 Firm-level wedge regressions

The firm-level survivorship wedge is the change in a survivor's percentile rank induced by conditioning, in percentile points:

$$w_{i,t} = 100 \left(\text{Pct}_{i,t}^S - \text{Pct}_{i,t}^F \right), \quad i \in S_t. \quad (12)$$

Because both percentiles are functions of $R_{i,t}$ alone, $w_{i,t} = g_t(R_{i,t})$ for a date-specific function g_t . The return regressions are

$$r_{i,t \rightarrow t+H} = \gamma_t + \beta w_{i,t} + m(R_{i,t}; \theta) + \varepsilon_{i,t}, \quad (13)$$

where $r_{i,t \rightarrow t+H}$ is the cumulative forward return ($H = 12$ months unless stated) and $m(\cdot)$ is the trailing-return control. Appendix Table A4 varies $m(\cdot)$ from none, to linear in $R_{i,t}$ or $\text{Pct}_{i,t}^F$, to cubic and quintic polynomials in the within-date standardized percentile, to date $\times$ percentile-bin fixed effects $\gamma_{t,q(i)}$ with 5, 10, 20, and 50 bins. As $m(\cdot)$ approaches a fully flexible function of $R_{i,t}$ within date, β is no longer identified. The state-dependent specification is

$$r_{i,t \rightarrow t+H} = \gamma_t + \beta_1 w_{i,t} + \beta_2 w_{i,t} Z_t + \theta_1 R_{i,t} + \theta_2 R_{i,t} Z_t + \varepsilon_{i,t}, \quad (14)$$

where Z_t is a standardized market-state proxy: the trailing twelve-month market return, or a composite of six standardized proxies. The main effect of Z_t is absorbed by γ_t , and the term $\theta_2 R_{i,t} Z_t$ ensures that β_2 is not state-dependent momentum. Appendix Table A5 estimates (14) for every combination of survival horizon $h \in \{12, 36, 60 \text{ months, end of sample}\}$, forward horizon $H \in \{12, 36, 60\}$, and proxy (24 specifications), and adjusts the p -values on β_2 by Benjamini and Hochberg (1995):

$$p_{(j)}^{\text{BH}} = \min_{l \geq j} \min \left(1, \frac{J}{l} p_{(l)} \right), \quad J = 24, \quad (15)$$

where $p_{(1)} \leq \dots \leq p_{(J)}$ are the ordered unadjusted p -values. Heterogeneity (Appendix Table A6, Panel B) adds a characteristic indicator $D_{i,t}$, defined relative to the within-date median, and its interaction:

$$r_{i,t \rightarrow t+12} = \gamma_t + \beta_1 w_{i,t} + \phi D_{i,t} + \psi w_{i,t} D_{i,t} + \theta_1 R_{i,t} + \varepsilon_{i,t}. \quad (16)$$

The two placebos in Appendix Table A6, Panel A, keep the right-hand side of (14), first without and then with the θ terms. The first replaces the dependent variable with the return over months -24 to -13 , which enters neither ranking,

$$r_{i,t-24 \rightarrow t-13} = \frac{1 + R_{i,t}^{(24)}}{1 + R_{i,t}} - 1, \quad (17)$$

where $R_{i,t}^{(24)}$ is the trailing 24-month return. The second replaces the wedge with a pseudo-wedge built from a future event that is not survival,

$$\tilde{w}_{i,t} = 100 \left(\widetilde{\text{Pct}}_{i,t}^S - \text{Pct}_{i,t}^F \right), \quad \tilde{S}_t = \{ i : \sigma_{i,t}^{\text{fwd}} \leq \text{median}_t(\sigma^{\text{fwd}}) \}, \quad (18)$$

where $\sigma_{i,t}^{\text{fwd}}$ is the annualized standard deviation of monthly returns over the twelve months after formation and $\widetilde{\text{Pct}}_{i,t}^S$ is the percentile rank of $R_{i,t}$ within $\tilde{S}_t$.

Standard errors for (11)–(18) are two-way clustered by security and formation date (Cameron, Gelbach, and Miller 2011; Petersen 2009), $\hat{V} = \hat{V}_{\text{security}} + \hat{V}_{\text{date}} - \hat{V}_{\text{security} \times \text{date}}$, with any negative eigenvalues set to zero. Because a security appears at most once per date, the last term is the heteroskedasticity-robust estimator. With 30 to 34 date clusters these standard errors can be unreliable, which is one reason the paper’s main inference is at the portfolio and date level, (3) and (6), rather than in the firm-level panel.

5 How Much Survivorship Reclassifies Winners

Table 1, Panel A, reports the share of top- K positions that change when the ranking is conditioned on survival. Reclassification increases with the survival horizon, from about 6 percent at one year to 14–19 percent at three years, 29–33 percent at five years, and 57–64 percent by the end of the sample. The pattern is similar across values of K . By Proposition 1, these rates are also the attrition rates among the original top- K firms.

Panel B compares the observed five-year reclassification rate with a benchmark that randomly removes the observed number of firms from each formation-date universe before re-ranking. Among the top twenty winners, 32.2 percent exit within five years, compared with 32.5 percent for the full universe. The permutation benchmark gives a mean reclassification rate of 32.4 percent (standard deviation 1.9 percentage points), with a one-sided p -value of 0.58 for the observed rate. The corresponding p -values are 0.75 for $K = 10$ and 0.68 and 0.82 when the random removal is additionally stratified by GICS sector.

Thus, the survivor-conditioned ranking changes the winner list substantially, but the amount of reclassification is similar to what would arise from random removal at the observed attrition rate. The distinction becomes important when exit type is considered: the next sections show that the outcomes associated with those exits are not similar.

Table 1. Winner Reclassification under Survivorship Conditioning

<i>Panel A: Share of top-K slots reclassified (%)</i>							
Survival horizon	$K = 1$	$K = 5$	$K = 10$	$K = 20$			
1 year	8.8	5.9	5.6	5.9			
3 years	18.8	13.8	18.1	18.9			
5 years	33.3	28.7	30.7	32.2			
End of sample	57.1	61.1	62.9	63.7			

<i>Panel B: Observed reclassification against random attrition (five-year horizon, %)</i>							
Strata	K	Observed	Universe attrition	Random-attrition benchmark			
				Mean	SD	95% range	p -value
Formation date	10	30.7	32.5	32.5	2.8	[27.3, 37.7]	0.75
Formation date	20	32.2	32.5	32.4	1.9	[28.7, 36.2]	0.58
Date $\times$ GICS sector	10	30.7	32.5	33.1	2.8	[27.3, 38.7]	0.82
Date $\times$ GICS sector	20	32.2	32.5	33.0	1.9	[29.3, 36.7]	0.68

Notes: Specifications: equations (2) and (3). Panel A reports the share of the K survivor-conditioned winner slots occupied by firms outside the full-ranking top K , averaged over formation dates (34, 32, 30, and 35 dates for the four horizons). By Proposition 1 this equals the attrition rate among full-ranking top- K firms over the horizon. Panel B compares the five-year rate with a null in which the observed number of exits is assigned at random within each stratum, the survivors are re-ranked, and the reclassification rate is recomputed; 1,000 replications. “Universe attrition” is the five-year exit rate of all firms, averaged over dates. The p -value is the share of replications with a rate at least as large as observed.

6 The Survivorship Premium and Its Decomposition

6.1 Portfolio returns

Table 2, Panel A, reports mean cumulative returns for the four equal-weighted winner portfolios, and Figure 1 shows their evolution after formation. The full-ranking portfolio earns 0.9 percent over twelve months and 21.4 percent over sixty months. The survivor-conditioned portfolio earns 7.1 and 58.8 percent, respectively. The difference is 6.2 percentage points at twelve months, 15.3 points at thirty-six months, and 37.4 points at sixty months. Panel C reports factor-adjusted estimates. The annualized alpha of the return difference is 5.2 percent under the CAPM and 5.0 percent under the Fama-French five-factor model augmented with momentum; the corresponding estimate is 4.5 percent under the five-factor model without momentum. These estimates indicate that the portfolio difference is not accounted for by the factor exposures included in these models.

The promoted firms and common winners provide a direct comparison because both groups survive to the end of the relevant horizon. Promoted firms have lower cumulative returns at most horizons beyond six months, including a difference of -4.3 percentage points at twelve months and -12.4 points at thirty-six months, but the date-level differences are not statistically distinguishable from zero. The factor-adjusted difference is -3.3 percent per year.

6.2 Decomposition

Table 3 applies equation (1) at each horizon. The truncation component is 8.3 percentage points at twelve months, 18.4 points at thirty-six months, and 34.7 points at sixty months. The composition component ranges from -3.1 to 2.7 percentage points and is not statistically distinguishable from zero in the pooled sample. Because the composition term is sometimes negative, the truncation component is larger than the total difference at some horizons. Figure 2 shows that the total difference closely follows the truncation component through the five-year period.

Panel B characterizes the firms underlying the truncation term. On average, 6.4 of the 20 full-ranking winners exit within five years, after 30 months. Of the 193 exits, 87 are distress delistings and 100 are acquisitions. Distress exits have a median return of -97 percent from formation to delisting, whereas acquired firms have a median return of 23 percent. While they remain listed, exiting winners earn -2.02 percent per month, compared with 0.92 percent for common winners in the same months; the difference is -2.95 percent per month. The survivor-conditioned portfolio excludes these observations, whereas the promoted firms do not generate a comparable return difference.

Table 2. Winner-Portfolio Returns

	Holding period (months)							
	1	3	6	12	36	60		
<i>Panel A: Mean cumulative return (%)</i>								
Full-ranking winners (W^F)	−1.03	−4.23	−0.38	0.90	4.03	21.43		
Survivor-conditioned winners (W^S)	−0.07	−1.55	2.71	7.11	19.28	58.84		
Common winners (C)	0.03	−2.07	2.63	9.23	22.42	56.14		
Promoted firms (P)	−0.59	0.02	3.87	4.92	9.97	51.08		
<i>Panel B: Return differences (percentage points)</i>								
$W^S - W^F$	0.96 (1.31)	2.67 (2.52)	3.09 (1.35)	6.21 (2.39)	15.25 (2.87)	37.42 (3.29)		
$P - C$	−0.62 (−0.50)	2.09 (0.74)	1.24 (0.27)	−4.32 (−0.72)	−12.45 (−1.57)	−5.05 (−0.39)		
<i>Panel C: Annualized alpha of return differences, 12-month holding periods (%)</i>								
	CAPM		FF3		FF5		FF5 + MOM	
	α	t	α	t	α	t	α	t
$W^S - W^F$	5.18	2.31	5.27	2.31	4.51	1.96	4.98	2.17
$P - C$	−2.10	−0.34	−1.66	−0.26	−4.21	−0.65	−3.32	−0.51

Notes: Specifications: Panels A and B, equations (4) and (6); Panel C, equation (7). Equal-weighted portfolios of the top 20 firms by trailing 12-month return, formed annually in June 1991–2020 (30 dates) using the five-year survival horizon. Constituents are rebalanced monthly across listed firms; a firm that delists contributes its return through its final month. Panels A and B average cumulative returns across formation dates; t -statistics in parentheses use one observation per date. Panel C regresses the calendar-time monthly return difference (360 months, from non-overlapping 12-month holding periods) on the indicated factors, with Newey-West standard errors (four lags); alphas are annualized by compounding.

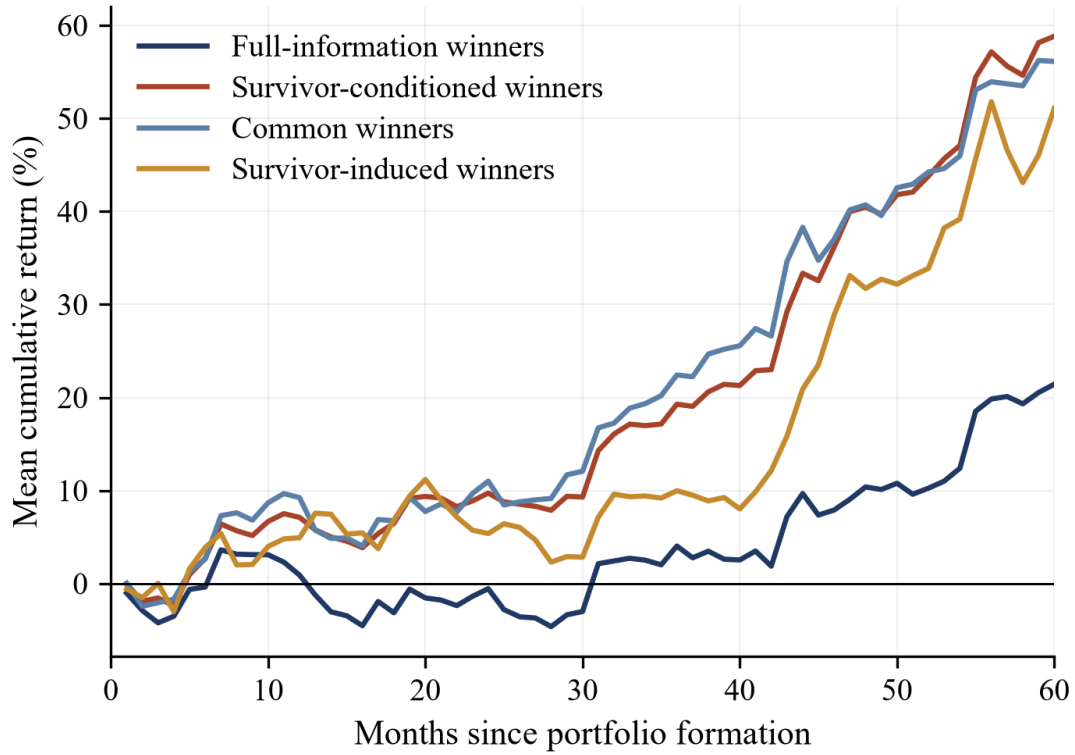


Figure 1. Event-Time Cumulative Returns of Winner Portfolios

Notes: Mean cumulative equal-weighted return by months since formation, averaged over 30 formation dates (1991–2020), for the full-ranking winners, survivor-conditioned winners, common winners, and promoted (survivor-induced) firms defined in Section 2; $K = 20$, five-year survival horizon.

Table 3. Decomposing the Survivorship Premium*Panel A: Cumulative return gap and components (percentage points)*

Holding period	Total gap $R^S - R^F$	Composition $R^S - R^C$	Truncation $R^C - R^F$	Truncation share (%)
1 month	0.96 (1.31)	-0.10 (-0.22)	1.05 (1.38)	110
3 months	2.67 (2.52)	0.51 (0.58)	2.16 (1.66)	81
6 months	3.09 (1.35)	0.07 (0.05)	3.01 (1.65)	98
12 months	6.21 (2.39)	-2.13 (-1.00)	8.33 (2.67)	134
24 months	10.28 (3.14)	-1.28 (-0.51)	11.56 (3.20)	112
36 months	15.25 (2.87)	-3.14 (-1.01)	18.39 (3.78)	121
48 months	30.08 (2.64)	-0.25 (-0.04)	30.33 (3.74)	101
60 months	37.42 (3.29)	2.71 (0.55)	34.71 (3.73)	93

Panel B: Exiting winners (E)

Exiting winners per formation date (of 20)	6.43
Total exits (30 dates)	193
Distress delistings / acquisitions / voluntary	87 / 100 / 6
Mean months from formation to exit: all / distress / acquisition	29.8 / 33.4 / 26.5
Return from formation to exit, distress exits: mean / median (%)	-85.8 / -97.1
Return from formation to exit, acquired firms: mean / median (%)	62.0 / 23.3
Mean monthly return while listed: exiting winners (%)	-2.02
Mean monthly return in the same months: common winners (%)	0.92
Difference (%), t -statistic	-2.95 (-5.89)

Notes: Specifications: Panel A, equations (5) and (6); Panel B, equations (8) and (9). Panel A applies the identity $R^S - R^F = (R^S - R^C) + (R^C - R^F)$ to the equal-weighted cumulative returns in Table 2 (30 formation dates, $K = 20$, five-year survival horizon). t -statistics in parentheses use one observation per formation date. Truncation share is the truncation component divided by the total gap; it exceeds 100 percent when the composition component is negative. In Panel B, returns from formation to exit are cumulative CRSP returns through the final month of listing. The monthly return of exiting winners is recovered exactly from the full and common portfolios, $r_\tau^E = (n_\tau^F r_\tau^F - n_\tau^C r_\tau^C) / (n_\tau^F - n_\tau^C)$, where n is the number of listed constituents; the t -statistic uses the per-date average difference.

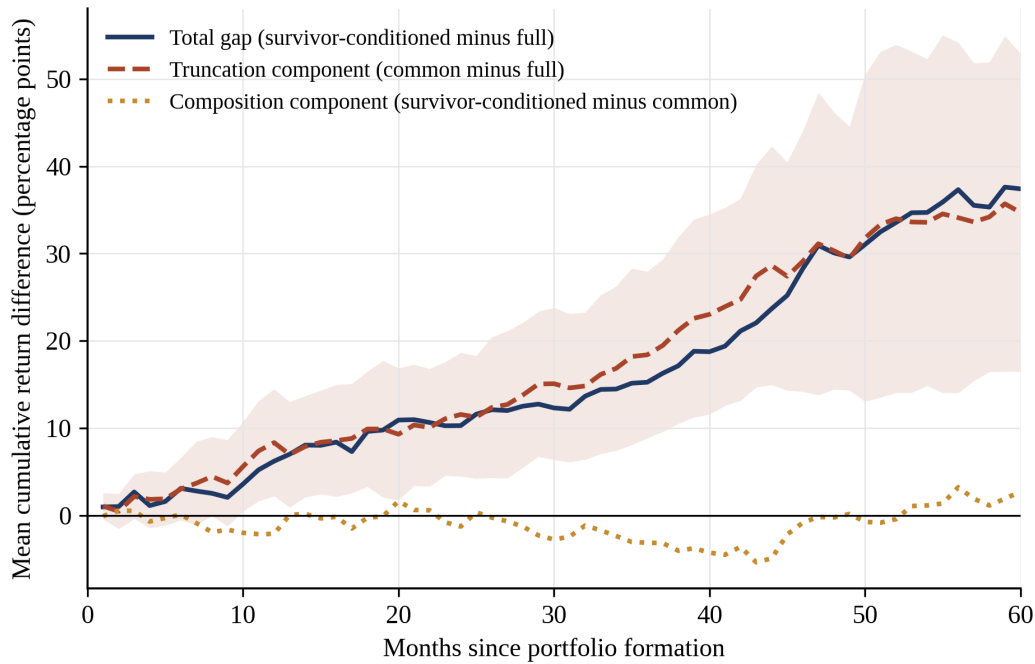


Figure 2. Decomposition of the Survivorship Premium in Event Time

Notes: Mean difference in cumulative equal-weighted return between winner portfolios, by months since formation, over 30 formation dates (1991–2020). The total gap is the survivor-conditioned minus the full-ranking portfolio; the truncation component is the common-winner minus the full-ranking portfolio; the composition component is the survivor-conditioned minus the common-winner portfolio. The two components sum to the total. The shaded band is a 95 percent confidence interval for the truncation component based on cross-date variation.

7 Why Truncation Bites in the Winner Tail

The reclassification result and the return decomposition can be reconciled by looking at the type of exit. Total exit rates need not differ between winners and the full universe for the associated return outcomes to differ.

Figure 3 plots five-year exit rates across the trailing-return distribution, and Table 4 reports selected groups. Total exit is U-shaped: 52.6 percent in the bottom decile, about 30 percent through the middle and top deciles, and 32.2 percent among the top twenty winners. Distress delistings are especially common in the bottom decile, decline through the middle of the distribution, and rise again in the upper tail. They reach 11.5 percent in the top percentile and 14.5 percent among the top twenty winners. Acquisition rates are comparatively stable through much of the distribution but fall to 16.7 percent among the top twenty winners.

Relative to the full universe, the top twenty winners have a distress-exit rate that is 5.2 percentage points higher and an acquisition rate that is 4.8 points lower. These differences nearly offset in total exit rates: the top-twenty rate is 0.3 points lower than the universe rate. The composition of exits nevertheless differs: distress accounts for 43.5 percent of winner exits versus 27.6 percent for the full universe. Since distress exits are associated with large losses while acquisitions tend to end with positive returns, the exit mix helps explain why removing winners can have a substantial effect on portfolio returns even when total attrition is similar.

The exit distribution also illustrates why estimates that impose a single linear relation between past returns and exit can be difficult to interpret. Over the full cross-section, the bottom tail strongly influences the estimated relation (Appendix Table A1), whereas the winner portfolios use only the extreme upper tail. The relevant exit pattern therefore depends on the portion of the return distribution being selected.

8 Robustness and Scope

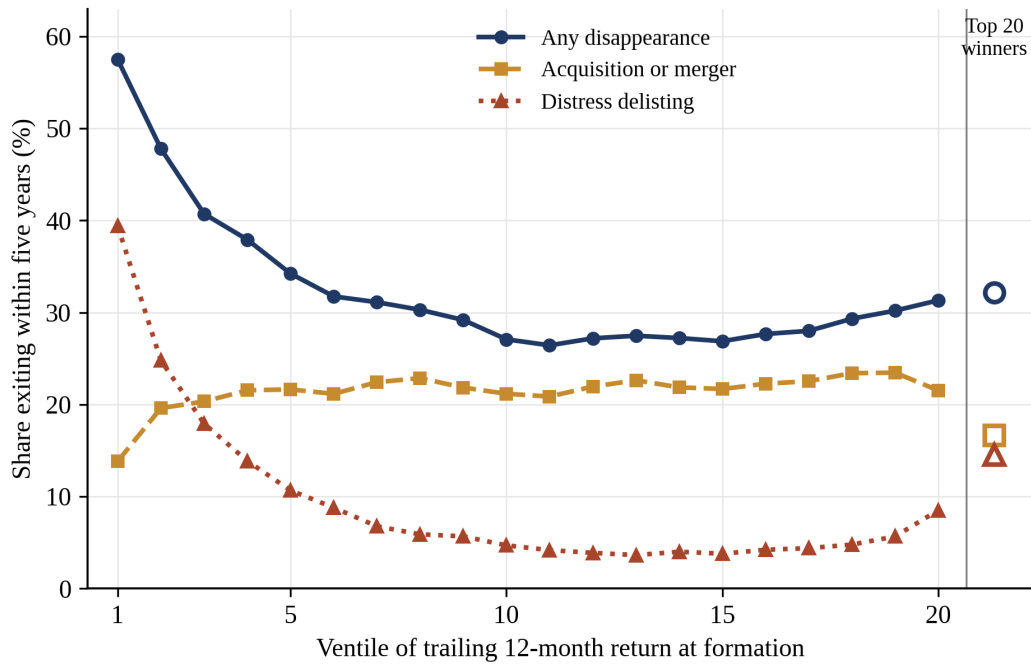
Subperiods, overlap, and weighting. the robustness analysis repeats the decomposition by subperiod and under value weighting. The total difference and truncation component are positive in both halves of the sample at the reported horizons, although estimates are less precise in the later period. For example, at thirty-six months the truncation component is 30.6 percentage points in 1991–2005 and 6.2 points in 2006–2020. The promoted-firm component is negative and statistically distinguishable from zero at some horizons in the earlier period, indicating that composition is not mechanically zero in every subsample. Across non-overlapping formation dates, the sign and broad magnitude of the equal-weighted difference remain similar, although the smaller number of observations widens the sampling uncertainty.

Value weighting produces a different pattern. The twelve-month difference is 12.4 percentage points, and none of the reported horizons is statistically different from zero at the 5 percent level. The point estimates are not driven by the truncation component in the same way as under equal weighting. This

Table 4. Five-Year Exit Rates across the Trailing-Return Distribution

Trailing-return group	N	Exit rate within five years (%)				Distress share of exits (%)
		Any	Distress	Acquisition	Voluntary	
All firms	146,052	32.5	9.3	21.4	1.7	27.6
Bottom decile	14,591	52.6 (19.86)	32.1 (20.32)	16.7 (−6.76)	3.8 (4.82)	60.1
Deciles 2–9	116,843	30.2 (−19.05)	6.7 (−21.42)	21.9 (3.80)	1.6 (−3.82)	21.2
Top decile	14,618	30.8 (−2.30)	7.1 (−4.46)	22.5 (1.80)	1.2 (−4.95)	22.2
Top percentile	1,478	31.5 (−0.66)	11.5 (2.04)	18.8 (−2.22)	1.2 (−1.72)	34.7
Top 20 winners	600	32.2 (−0.16)	14.5 (3.00)	16.7 (−3.07)	1.0 (−2.34)	43.5

Notes: Specification: equation (10), tested with (6). Five-year exit rates by group of trailing 12-month return at formation, 30 formation dates (1991–2020). Rates are computed within each formation date and averaged across dates; “Distress share of exits” is the date-averaged ratio of distress exits to all exits. *t*-statistics in parentheses test whether the group’s rate differs from the all-firm rate on the same date, using one paired difference per date. The pooled (observation-weighted) all-firm rates are 33.9, 10.1, 22.0, and 1.8 percent.

**Figure 3.** Exit Type across the Trailing-Return Distribution

Notes: Share of firms exiting within five years of formation, by ventile of trailing 12-month return and by exit type, averaged over 30 formation dates (1991–2020). Hollow markers at the right show the same rates for the top 20 firms by trailing return.

difference is consistent with the relatively small size of firms that later exit: their mean size percentile is 0.45 among all top-twenty winners and 0.39 among those exiting through distress, compared with 0.49 for the full winner group. The main estimates therefore describe the behavior of equal-weighted winner portfolios rather than a general effect across portfolio weighting schemes.

The firm-level wedge. The firm-level survivorship measure, defined as the difference between a survivor’s percentile rank under the survivor-conditioned and full rankings, varies systematically with the trailing-return distribution (Appendix Figure A2). Within a formation date it is determined by the firm’s own trailing return, so flexible controls for that return absorb most of the variation in the measure. Accordingly, the coefficient on the survivorship measure changes substantially as the trailing-return control becomes more flexible (Appendix Table A4). This exercise does not provide a separate source of return predictability beyond the information already contained in the formation-date return ranking.

Extrapolative market states. We also examine whether the firm-level measure varies with market conditions associated with extrapolative expectations. Interacting the measure with standardized market-state proxies while controlling for the firm’s trailing return and its interaction with the state variable yields one nominally significant interaction at the 5 percent level across 24 specifications, and none remains significant after Benjamini-Hochberg adjustment (Appendix Table A5). A placebo based on future realized volatility also produces a significant interaction in one specification, which further cautions against interpreting individual interaction estimates as evidence of state dependence. Given the number of formation dates, these results are more appropriately viewed as limited evidence on heterogeneity than as a precise test of state dependence.

Construction checks. The firm-level measure does not predict returns over months -24 to -13 , which are not used in the formation ranking (Appendix Table A6). This provides a check that the construction is not mechanically related to an earlier return window. Using a top-decile rather than a fixed- K winner definition produces fewer survivor-conditioned holdings (322 versus 487) and a higher value-weighted Herfindahl index (0.049 versus 0.041), as expected when the survivor-conditioned universe is smaller (Appendix Table A3).

9 Conclusion

Survivorship conditioning changes a ranking of past winners in two distinct ways. It can change which firms enter the winner set, and it can remove the subsequent returns of winners that later exit. Using point-in-time U.S. equity data, we find that the first channel is substantial in terms of the number of positions affected but is similar in magnitude to what overall attrition would imply. The second channel is more closely associated with the difference in portfolio returns: for an equal-weighted portfolio of

the top twenty past winners, the survivor-conditioned backtest exceeds the full-universe backtest by 6.2 percentage points after one year and 37.4 points after five years. The decomposition attributes most of this difference to the returns of winners that exit rather than to the firms promoted into their place.

The exit distribution helps account for the result. The top twenty winners do not have a materially different total five-year exit rate from the full universe, but their exits are more likely to be distress-related and less likely to be acquisitions. Distress exits are associated with large losses before delisting. A survivor-only database therefore changes the observed return distribution even when the number of omitted firms is not unusual.

The results have a direct implication for empirical work that evaluates winner-selection strategies. The amount of reclassification in a survivor-only database is not, by itself, a measure of the resulting performance difference. A database can replace many names simply because attrition is common, while still producing a substantial difference in measured performance because the omitted outcomes are concentrated among firms with poor subsequent returns. The magnitude of the effect also depends on portfolio weighting and sample period, so the estimates should not be interpreted as a universal adjustment to all survivorship-affected performance measures.

More broadly, the decomposition provides a way to distinguish changes in selected units from missing subsequent outcomes whenever a sample is formed using past performance and later conditioned on survival. Applying the same accounting to other settings, such as funds, managers, or private firms, would require data that preserve both the original selection set and subsequent exits. Such applications could help determine how much of survivorship bias in other performance-evaluation settings arises from changes in composition versus omitted outcomes.

References

- Aiken, A. L., C. P. Clifford, and J. A. Ellis. 2013. Out of the dark: Hedge fund reporting biases and commercial databases. *Review of Financial Studies* 26:208–243.
- Amihud, Y. 2002. Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets* 5:31–56.
- Avramov, D., T. Chordia, G. Jostova, and A. Philipov. 2007. Momentum and credit rating. *Journal of Finance* 62:2503–2520.
- Barberis, N., R. Greenwood, L. Jin, and A. Shleifer. 2015. X-CAPM: An extrapolative capital asset pricing model. *Journal of Financial Economics* 115:1–24.
- Barberis, N., R. Greenwood, L. Jin, and A. Shleifer. 2018. Extrapolation and bubbles. *Journal of Financial Economics* 129:203–227.
- Barberis, N., L. J. Jin, and B. Wang. 2021. Prospect theory and stock market anomalies. *Journal of Finance* 76:2639–2687.
- Beaver, W., M. McNichols, and R. Price. 2007. Delisting returns and their effect on accounting-based market anomalies. *Journal of Accounting and Economics* 43:341–368.
- Ben-David, I., J. Li, A. Rossi, and Y. Song. 2022. What do mutual fund investors really care about? *Review of Financial Studies* 35:1723–1774.
- Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B* 57:289–300.
- Bessembinder, H. 2018. Do stocks outperform Treasury bills? *Journal of Financial Economics* 129:440–457.
- Bordalo, P., N. Gennaioli, R. La Porta, and A. Shleifer. 2019. Diagnostic expectations and stock returns. *Journal of Finance* 74:2839–2874.
- Bordalo, P., N. Gennaioli, R. La Porta, and A. Shleifer. 2024. Belief overreaction and stock market puzzles. *Journal of Political Economy* 132:1450–1484.
- Brown, S. J., W. N. Goetzmann, R. G. Ibbotson, and S. A. Ross. 1992. Survivorship bias in performance studies. *Review of Financial Studies* 5:553–580.
- Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2011. Robust inference with multiway clustering. *Journal of Business & Economic Statistics* 29:238–249.
- Campbell, J. Y., J. Hilscher, and J. Szilagyi. 2008. In search of distress risk. *Journal of Finance* 63:2899–2939.
- Carhart, M. M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52:57–82.
- Carhart, M. M., J. N. Carpenter, A. W. Lynch, and D. K. Musto. 2002. Mutual fund survivorship. *Review of Financial Studies* 15:1439–1463.
- Carpenter, J. N., and A. W. Lynch. 1999. Survivorship bias and attrition effects in measures of performance persistence. *Journal of Financial Economics* 54:337–374.
- Chen, A. Y., and T. Zimmermann. 2022. Open source cross-sectional asset pricing. *Critical Finance Review* 11:207–264.
- Cochrane, J. H. 2005. The risk and return of venture capital. *Journal of Financial Economics* 75:3–52.
- Conrad, J., N. Kapadia, and Y. Xing. 2014. Death and jackpot: Why do individual investors hold overpriced stocks? *Journal of Financial Economics* 113:455–475.

- Da, Z., X. Huang, and L. J. Jin. 2021. Extrapolative beliefs in the cross-section: What can we learn from the crowds? *Journal of Financial Economics* 140:175–196.
- Daniel, K., and T. J. Moskowitz. 2016. Momentum crashes. *Journal of Financial Economics* 122:221–247.
- Elton, E. J., M. J. Gruber, and C. R. Blake. 1996. Survivorship bias and mutual fund performance. *Review of Financial Studies* 9:1097–1120.
- Evans, R. B. 2010. Mutual fund incubation. *Journal of Finance* 65:1581–1611.
- Evans, R. B., and Y. Sun. 2021. Models or stars: The role of asset pricing models and heuristics in investor risk adjustment. *Review of Financial Studies* 34:67–107.
- Fama, E. F., and K. R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33:3–56.
- Fama, E. F., and K. R. French. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116:1–22.
- Greenwood, R., and A. Shleifer. 2014. Expectations of returns and expected returns. *Review of Financial Studies* 27:714–746.
- Grinblatt, M., S. Titman, and R. Wermers. 1995. Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. *American Economic Review* 85:1088–1105.
- Harvey, C. R., and Y. Liu. 2020. False (and missed) discoveries in financial economics. *Journal of Finance* 75:2503–2553.
- Harvey, C. R., Y. Liu, and H. Zhu. 2016. ... and the cross-section of expected returns. *Review of Financial Studies* 29:5–68.
- Hendricks, D., J. Patel, and R. Zeckhauser. 1993. Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988. *Journal of Finance* 48:93–130.
- Hou, K., C. Xue, and L. Zhang. 2020. Replicating anomalies. *Review of Financial Studies* 33:2019–2133.
- Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48:65–91.
- Jensen, T. I., B. Kelly, and L. H. Pedersen. 2023. Is there a replication crisis in finance? *Journal of Finance* 78:2465–2518.
- Kaniel, R., and R. Parham. 2017. WSJ Category Kings – The impact of media attention on consumer and mutual fund investment decisions. *Journal of Financial Economics* 123:337–356.
- Korteweg, A., and M. Sorensen. 2010. Risk and return characteristics of venture capital-backed entrepreneurial companies. *Review of Financial Studies* 23:3738–3772.
- Kothari, S. P., J. Shanken, and R. G. Sloan. 1995. Another look at the cross-section of expected stock returns. *Journal of Finance* 50:185–224.
- Linnainmaa, J. T. 2013. Reverse survivorship bias. *Journal of Finance* 68:789–813.
- Linnainmaa, J. T., and M. R. Roberts. 2018. The history of the cross-section of stock returns. *Review of Financial Studies* 31:2606–2649.
- McLean, R. D., and J. Pontiff. 2016. Does academic research destroy stock return predictability? *Journal of Finance* 71:5–32.
- Menkveld, A. J., A. Dreber, F. Holzmeister, J. Huber, M. Johannesson, M. Kirchler, et al. 2024. Nonstandard errors. *Journal of Finance* 79:2339–2390.

- Newey, W. K., and K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–708.
- Petersen, M. A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22:435–480.
- Schwarz, P., D. Walter, and P. Weiss. 2026. Rewriting CRSP’s history: Impact of altered monthly returns on asset pricing. *Journal of Financial and Quantitative Analysis*, forthcoming (published online March 2026).
- Shumway, T. 1997. The delisting bias in CRSP data. *Journal of Finance* 52:327–340.
- Shumway, T., and V. A. Warther. 1999. The delisting bias in CRSP’s Nasdaq data and its implications for the size effect. *Journal of Finance* 54:2361–2379.
- van Binsbergen, J. H., S. Hua, J. Peeters, and J. A. Wachter. 2025. Is the United States a lucky survivor? A hierarchical Bayesian approach. *Journal of Finance* 80:2355–2388.

Appendix

A Proof of Proposition 1

Rank firms in descending order of R_i and break ties by a fixed rule common to both rankings. For $i \in S_t$, the set of survivors ranked above i is a subset of the firms in U_t ranked above i , so $\text{rank}_{S_t}(i) \leq \text{rank}_{U_t}(i)$. Hence every $i \in W_t^F \cap S_t = C_t$ has survivor rank at most K and belongs to W_t^S . Every other member of W_t^S is outside W_t^F by definition, so $W_t^S = C_t \cup P_t$ with $C_t \cap P_t = \emptyset$. Since $|S_t| \geq K$, $|W_t^S| = K = |C_t| + |E_t|$, which gives $|P_t| = K - |C_t| = |E_t|$. For the last statement, if the $|S_t|$ survivors are drawn uniformly at random from U_t , each firm exits with probability $1 - |S_t|/|U_t|$, so $E|E_t| = K(1 - |S_t|/|U_t|)$ and $E[\rho_t]$ equals the universe attrition rate.

B Supplementary Tables and Figures

Table A1. Linear Probability Models of Exit

	12 months			36 months			60 months		
	Any	Distr.	Acq.	Any	Distr.	Acq.	Any	Distr.	Acq.
Return percentile (0–100)	0.0002** (0.0001)	−0.0003*** (0.0000)	0.0004*** (0.0001)	−0.0001 (0.0001)	−0.0007*** (0.0001)	0.0006*** (0.0001)	−0.0002 (0.0001)	−0.0009*** (0.0001)	0.0007*** (0.0001)
Trailing 12-month return	−0.0093*** (0.0026)	−0.0056*** (0.0019)	−0.0039*** (0.0012)	−0.0111*** (0.0030)	−0.0081*** (0.0024)	−0.0040** (0.0019)	−0.0087** (0.0037)	−0.0076* (0.0038)	−0.0029 (0.0026)
Log size	−0.0069*** (0.0007)	−0.0037*** (0.0006)	−0.0012** (0.0005)	−0.0196*** (0.0019)	−0.0108*** (0.0014)	−0.0030* (0.0016)	−0.0286*** (0.0025)	−0.0165*** (0.0018)	−0.0031 (0.0024)
Log age	−0.0129*** (0.0011)	0.0017** (0.0007)	−0.0148*** (0.0011)	−0.0385*** (0.0026)	0.0014 (0.0016)	−0.0405*** (0.0025)	−0.0550*** (0.0035)	0.0007 (0.0023)	−0.0568*** (0.0036)
Idiosyncratic volatility	0.1201*** (0.0236)	0.1622*** (0.0218)	−0.0318*** (0.0113)	0.2756*** (0.0453)	0.4517*** (0.0491)	−0.1533*** (0.0327)	0.3228*** (0.0459)	0.6360*** (0.0586)	−0.2788*** (0.0465)
Leverage	0.0472*** (0.0050)	0.0439*** (0.0050)	0.0011 (0.0031)	0.1005*** (0.0113)	0.0964*** (0.0108)	−0.0013 (0.0088)	0.1171*** (0.0155)	0.1251*** (0.0138)	−0.0159 (0.0120)
Profitability	−0.0255*** (0.0039)	−0.0237*** (0.0035)	−0.0003 (0.0011)	−0.0403*** (0.0057)	−0.0429*** (0.0056)	0.0055** (0.0021)	−0.0375*** (0.0063)	−0.0460*** (0.0062)	0.0111*** (0.0028)
Base rate	0.079	0.026	0.049	0.223	0.069	0.143	0.339	0.101	0.220
Formation dates	34	34	34	32	32	32	30	30	30
Observations	133,699	133,699	133,699	126,900	126,900	126,900	120,549	120,549	120,549
Within R^2	0.032	0.087	0.005	0.058	0.134	0.011	0.071	0.149	0.015

Notes: Specification: equation (11). Linear probability models of exit within the indicated horizon, full point-in-time universe, with formation-date fixed effects. In addition to the reported regressors, every column controls for log book-to-market, 36-month market beta, and Amihud (2002) illiquidity; their coefficients are omitted to save space. Standard errors in parentheses are two-way clustered by security and formation date. Each horizon uses only formation dates for which it is fully observed. The percentile coefficient multiplied by 100 is the change in exit probability from the bottom to the top of the trailing-return distribution, holding the other regressors fixed; at 60 months it implies −9 percentage points for distress and 7 for acquisition. These linear fits are dominated by the bottom of the distribution and do not describe the extreme upper tail, where distress exits are more frequent than average (Table 4). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A2. Factor-Adjusted Alphas by Winner Portfolio

Portfolio	CAPM		FF3		FF5		FF5 + MOM	
	α	t	α	t	α	t	α	t
<i>Equal-weighted</i>								
Full-ranking winners	-14.4	-3.16	-13.6	-3.66	-9.5	-2.58	-9.2	-2.54
Survivor-conditioned winners	-9.9	-2.30	-9.0	-2.55	-5.3	-1.40	-4.7	-1.27
Common winners	-8.5	-1.69	-7.7	-1.76	-2.9	-0.61	-2.4	-0.52
Promoted firms	-10.4	-1.88	-9.2	-1.78	-7.0	-1.22	-5.6	-0.98
<i>Value-weighted</i>								
Full-ranking winners	-3.5	-0.54	-1.6	-0.27	2.6	0.40	1.3	0.20
Survivor-conditioned winners	3.9	0.57	6.2	0.97	7.4	1.07	5.6	0.84
Common winners	-3.5	-0.49	-1.5	-0.24	1.2	0.15	-0.4	-0.06
Promoted firms	4.5	0.56	6.9	0.88	11.3	1.33	10.7	1.25

Notes: Specification: equation (7), with the portfolio's excess return as the dependent variable. Annualized alpha (percent) and t -statistic from time-series regressions of each portfolio's monthly excess return on the indicated factors; 360 months from non-overlapping 12-month holding periods; Newey-West standard errors with four lags. The large negative equal-weighted alpha of the full-ranking portfolio reflects the losses of winners that later exit (Table 3).

Table A3. Portfolio Concentration by Winner Definition

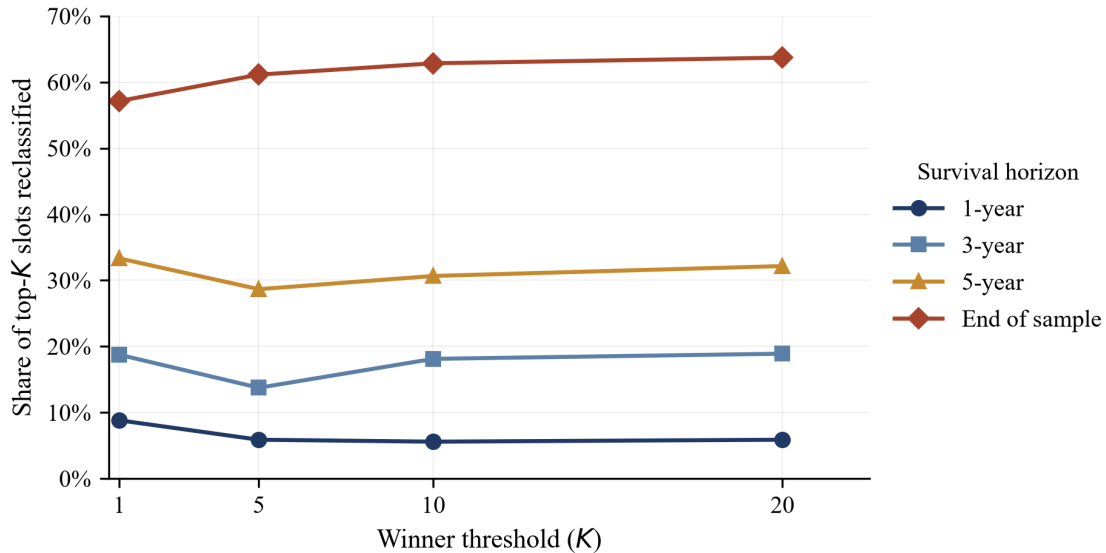
Definition	Ranking	Holdings	HHI (VW)	Effective N (VW)	Sectors
Top 10	Full	10.0	0.268	4.3	4.8
	Survivor-conditioned	10.0	0.284	4.1	4.6
Top 20	Full	20.0	0.194	6.8	6.3
	Survivor-conditioned	20.0	0.217	6.4	6.2
Top 50	Full	50.0	0.133	11.5	8.3
	Survivor-conditioned	50.0	0.139	11.5	8.3
Top decile	Full	487.4	0.041	39.1	10.9
	Survivor-conditioned	322.5	0.049	28.4	10.8

Notes: Averages over 30 formation dates. HHI and effective number of holdings use value weights; "Sectors" is the mean number of distinct GICS sectors. Under fixed K the equal-weighted HHI is identical across rankings by construction. Paired differences in value-weighted HHI (survivor-conditioned minus full) have t -statistics of 0.77, 1.26, and 0.29 for $K = 10, 20$, and 50, and 2.52 for the top decile.

Table A4. The Firm-Level Wedge and Momentum Controls

	Wedge variance surviving (%)	Wedge		Wedge $\times$ market state	
Trailing-return control		Coef.	t	Coef.	t
None	100.0	0.0017	0.32	−0.0012	−0.30
Linear trailing return	71.3	0.0171	3.92	−0.0013	−0.41
Linear trailing-return percentile	46.3	0.0367	3.00	0.0024	0.61
Cubic in percentile	14.1	0.0177	1.20	0.0013	0.36
Quintic in percentile	13.5	0.0142	0.87	0.0012	0.32
Date $\times$ quintile fixed effects	15.4	0.0430	3.19	−0.0098	−0.59
Date $\times$ decile fixed effects	5.7	0.0522	2.69	−0.0180	−0.75
Date $\times$ ventile fixed effects	1.9	0.0539	1.83	−0.0476	−1.52
Date $\times$ 50-tile fixed effects	0.4	−0.0003	−0.01	−0.0941	−2.66

Notes: Specifications: equation (13) in the wedge columns; in the interaction columns, equation (13) augmented with $w_{i,t}Z_t$ and $R_{i,t}Z_t$. Dependent variable is the forward 12-month return. The wedge is the difference between a firm's survivor-conditioned and full-universe trailing-return percentile (percentage points, five-year survival horizon); the market state is the standardized trailing 12-month market return. "Wedge variance surviving" is the share of the wedge's within-date variance not explained by the control. All specifications include formation-date fixed effects and, in the interaction columns, the trailing return interacted with the market state; standard errors are two-way clustered by security and formation date; 96,575 observations, 30 dates. As the control approaches a fully flexible function of the trailing return, the wedge's identifying variation vanishes.

**Figure A1.** Winner Reclassification by Threshold and Survival Horizon

Notes: Share of the K survivor-conditioned winner slots occupied by firms outside the full-ranking top K , averaged over formation dates, by survival horizon. By Proposition 1 each point is also the attrition rate among the full-ranking top- K firms. Values are reported in Table 1, Panel A.

Table A5. Extrapolative Market States: Full Specification Grid

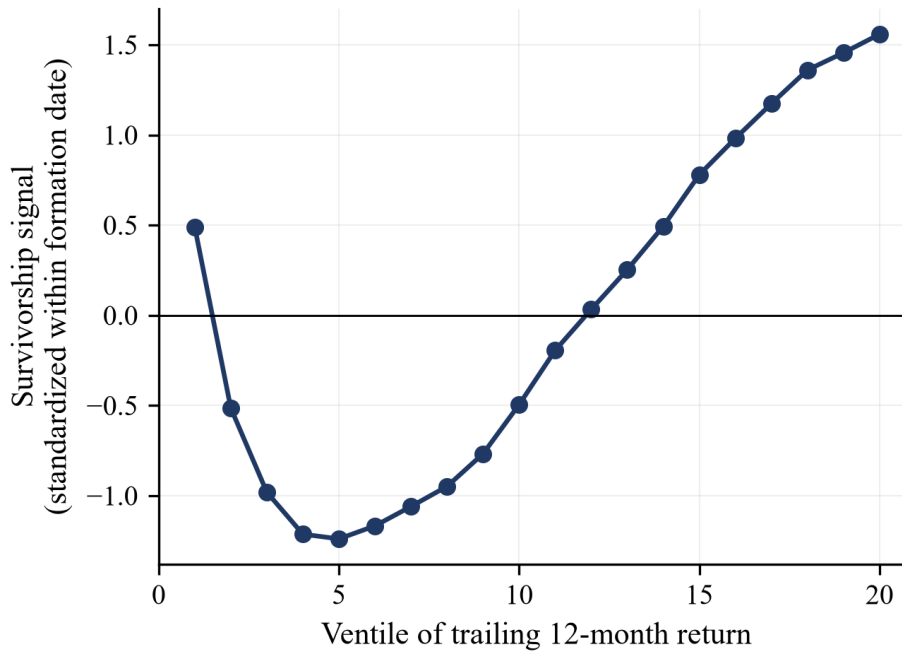
Survival horizon	Forward return	Market-state proxy	Coef.	<i>t</i>	<i>p</i>	BH <i>p</i>
End of sample	60 months	Composite	−0.0243	−2.82	0.009	0.204
End of sample	36 months	Composite	−0.0103	−1.98	0.057	0.497
1 year	60 months	Market return	0.0412	1.88	0.070	0.497
3 years	36 months	Composite	−0.0328	−1.79	0.083	0.497
3 years	36 months	Market return	−0.0097	−1.53	0.136	0.614
5 years	36 months	Market return	−0.0072	−1.46	0.156	0.614
End of sample	60 months	Market return	−0.0132	−1.19	0.244	0.614
3 years	60 months	Composite	−0.0420	−1.14	0.262	0.614
5 years	36 months	Composite	−0.0172	−1.13	0.268	0.614
5 years	60 months	Composite	−0.0249	−1.13	0.269	0.614
5 years	12 months	Composite	0.0145	1.05	0.304	0.614
1 year	12 months	Composite	−0.0330	−1.02	0.313	0.614
1 year	36 months	Composite	−0.0466	−0.98	0.332	0.614
1 year	12 months	Market return	0.0092	0.75	0.459	0.786
5 years	60 months	Market return	−0.0061	−0.54	0.595	0.876
3 years	12 months	Composite	0.0080	0.44	0.664	0.876
3 years	12 months	Market return	0.0018	0.42	0.674	0.876
5 years	12 months	Market return	−0.0013	−0.41	0.685	0.876
1 year	60 months	Composite	−0.0344	−0.35	0.727	0.876
End of sample	36 months	Market return	−0.0008	−0.35	0.730	0.876
3 years	60 months	Market return	−0.0019	−0.17	0.870	0.977
End of sample	12 months	Market return	−0.0001	−0.06	0.951	0.977
1 year	36 months	Market return	0.0008	0.05	0.964	0.977
End of sample	12 months	Composite	−0.0003	−0.03	0.977	0.977

Notes: Specification: equation (14); adjusted *p*-values from equation (15). Each row reports the coefficient on the interaction between the firm-level survivorship wedge and a standardized market-state proxy in a regression of the forward return, controlling for the firm’s trailing return and its interaction with the same proxy, with formation-date fixed effects and standard errors two-way clustered by security and formation date. “Market return” is the standardized trailing 12-month market return; “Composite” averages six standardized proxies (trailing return, valuation trend, breadth, dispersion, and two concentration measures). BH *p* is the Benjamini-Hochberg adjusted *p*-value across all 24 specifications; none is below 0.10. The primary pre-specified cell is the five-year horizon, 12-month forward return, and market-return proxy.

Table A6. Placebo Tests and Heterogeneity

		Wedge		Interaction	
	<i>N</i>	Coef.	<i>t</i>	Coef.	<i>t</i>
<i>Panel A: Placebo tests</i>					
Pre-formation return (months −24 to −13)	92,812	−0.0022	−0.42	0.0091	1.21
with trailing-return controls	92,812	0.0026	0.68	0.0091	1.99
Pseudo-survival from future volatility	80,398	0.0013	1.73	0.0016	2.15
with trailing-return controls	80,398	0.0007	1.03	0.0024	3.97
<i>Panel B: Interaction of the wedge with firm characteristics</i>					
Young firms	96,575	0.0148	2.97	0.0048	1.04
Small firms	96,575	0.0222	3.04	−0.0092	−0.87
Illiquid firms	96,574	0.0212	3.38	−0.0079	−0.90
High idiosyncratic volatility	86,114	0.0162	1.79	0.0011	0.07
High fundamental mismatch	93,856	0.0254	3.00	−0.0199	−1.54
High book-to-market	91,359	0.0225	3.77	−0.0068	−0.66

Notes: Specifications: Panel A, equation (14) with the dependent variable or the wedge replaced as in equations (17) and (18); Panel B, equation (16). Panel A: the pre-formation placebo regresses the return over months −24 to −13, which enters neither ranking, on the wedge; the interaction is with the market-state proxy. The pseudo-survival placebo rebuilds the wedge by treating low realized forward volatility, rather than survival, as the conditioning event; a significant interaction there indicates that the construction can produce interaction estimates from generic future information. Panel B: each row adds an indicator for the named characteristic (below the within-date median of age or size for young and small firms, above the within-date median otherwise) and its interaction with the wedge, controlling for the trailing return; dependent variable is the forward 12-month return. Formation-date fixed effects and two-way clustered standard errors throughout.

**Figure A2.** The Firm-Level Survivorship Wedge across the Trailing-Return Distribution

Notes: Mean difference between a firm's survivor-conditioned and full-universe percentile rank (five-year horizon), standardized within formation date, by ventile of trailing 12-month return; 30 formation dates.