CREDIT MISPRICING IN JAPAN:

REPURPOSING THE ZOMBIE-FIRM LITERATURE FOR THE NIKKEI 225

Andreas Grünbichler

Samu Elinger

Working Paper (SSRN abstract no. 6813479)

UNIVERSITY OF ST. GALLEN

St. Gallen, Switzerland

Current draft August 2026

v5.8

We thank Naomi Fink (Amova Asset Management), Rolf Banz, Takeo Hoshi (University of Tokyo), Takeshi Osada (Saitama University), Günter Strobl (University of Vienna), Takao Suzuki (Amova Asset Management), Masayuki Teraguchi (Amova Asset Management), and seminar participants at Saitama University for very valuable comments and suggestions. Jonatan Wimmer, our data scientist, provided invaluable support in the end-to-end data-quality verification underlying this revision. Without the guidance of Joji Hattori, Rick Dyck, Max Weber, and Hirotaka Hideshima on Japanese institutional context and the support of Hiroshi Nakaso, this paper would not have been possible. All remaining errors are our own. The views expressed here do not necessarily reflect those of any institution mentioned.

Credit Mispricing in Japan:
Repurposing the Zombie-Firm Literature for the Nikkei 225

Andreas Grünbichler and Samu Elinger

Current draft August 2026 (v5.8)

ABSTRACT

The academic zombie-firm literature counts distressed firms and measures their macroeconomic costs. It pays far less attention to what happens next. On the Tokyo Stock Exchange, classified firms rarely exit, and most recover. This paper turns that recovery gap into an investment strategy. We convert ten published zombie classifications into continuous credit-quality scores and form equally weighted portfolios of the most zombie-like decile of the Nikkei 225, rebalanced annually under a strict point-in-time protocol. Over 2000–2025, the three headline classifications deliver compound annual returns of 7.2 to 10.2 percent against 3.8 percent for the benchmark, net of a 50 basis point transaction-cost charge. Spanning tests against identically constructed value portfolios isolate the credit component: the coverage-based classifications retain 2.3 to 3.0 percentage points per annum over the value placebos, with a joint estimate of 2.7 points (t = 1.8). This premium is economically meaningful, statistically moderate on its own, and doubly concentrated: it sits precisely where classified firms show no elevated exit risk, and it is confined to the bank-connected blue-chip segment. On the broad TOPIX universe the same deciles are fully spanned by generic value. Both patterns are the signature of credit mispricing rather than compensation for default. The signal remains investable even in years when binary zombie counts are low.

JEL classification: G11, G12, G14, G32.
Keywords: zombie firms, credit mispricing, value investing, Japanese equities, TOPIX, Nikkei 225, Carhart factors, Schumpeter.

1. Introduction

The literature on zombie firms has grown rapidly since Caballero, Hoshi and Kashyap (2008) coined the term in their study of post-bubble Japan. Subsequent contributions — Adalet McGowan, Andrews and Millot (2017), Banerjee and Hofmann (2018, 2020), Acharya, Crosignani, Eisert and Eufinger (2020), Favara, Minoiu and Perez-Orive (2021), Altman, Dai and Wang (2022), and Nakamura (2023), among others — have refined the definition, mapped the geography of zombification, and quantified the macroeconomic costs of impaired creative destruction. With very few exceptions, the literature has been read as a diagnostic instrument: it tells policymakers and supervisors how many firms in the corporate sector are impaired, in which industries they cluster, and with what consequences for productivity and competition. It has not been read as a source of cross-sectional information about expected equity returns. This paper makes that bridge. We argue that the same balance-sheet ratios that the zombie-firm literature uses to diagnose distress (interest coverage ratio (ICR), Tobin's Q, leverage, real sales growth, and Altman's Z- and Z″-Scores) also contain investable information about the cross-section of expected returns in liquid, exchange-listed equities. The economic mechanism is credit mispricing. Equity investors apply substantial risk discounts to firms that look like zombies on standard accounting metrics, often well in excess of the realized rate at which those firms default. The realized recovery rate of classified zombies, in turn, is materially higher than market pricing implies. A systematic long position in firms with elevated zombie indicators therefore captures a risk-premium gap that compensates investors for credit risk that, on average, does not materialize. Two design choices distinguish our work from prior empirical studies. First, we replace the binary zombie flags that dominate the literature with continuous fuzzy scores. The binary flag, while appropriate for the macro question of how much of the corporate sector is impaired, is mechanically unsuitable for portfolio construction: it discards severity information and it does not produce a ranking. The fuzzy transformation produces a continuous, dimensionless, cross-sectional score that is well-defined for every Nikkei 225 constituent in every year. Second, we operate the score on the actual time-varying constituency of the Nikkei 225, with all signals constructed from information available at the rebalancing date. The backtest is therefore free of survivorship bias and look-ahead bias by construction; it is a point-in-time backtest in the sense defined in Section 5.2. A central practical implication of the fuzzy construction is what we call cycle-robustness. The binary literature partitions a sample into zombies and non-zombies, and the count of binary zombies fluctuates with the macroeconomic environment: it was high in Japan in the 1990s and early 2000s and has been low since the mid-2010s. A binary strategy is therefore mechanically dependent on the prevalence of zombification in the sample. The fuzzy score, by contrast, ranks the entire cross-section on cardinal credit-quality information every year. In years with many classical zombies the score concentrates the top decile in obviously impaired names. In years with few classical zombies the score still identifies the firms that are relatively most credit-impaired within the index, and the cross-sectional ranking remains well-defined. The signal is therefore as relevant in the 2020s as it was during the lost decades.

We highlight a related observation about the zombie-firm literature itself. The literature has been preoccupied with the count and the consequences of zombification; it has paid surprisingly limited

attention to what happens to classified firms after the classification date. How many of them default? How many recover? On what time horizon? With what fundamental drivers? Section 3 of this paper presents Kaplan–Meier survival analysis (Kaplan and Meier 1958) and Cox proportional-hazards regressions (Cox 1972) on a three-decade panel of the Tokyo Stock Exchange that begin to answer those questions. We document, first, that the median duration of an unbroken zombie spell is one to two years across the classifications; second, that annual exit intensities from the zombie pool are at most 1.7% (against roughly 0.6% for never-classified firms) and that annual relapse probabilities range from approximately 1% to 5.5%; and third, that the conditional probability of exit from the zombie pool is materially reduced by capital intensity (the classical collateral channel) and by liquidity, while the associations with sector market share, leverage, and asset productivity are close to flat; exits in the panel include takeovers. Recovery, not exit, is the modal outcome. That asymmetry is the foundation of the credit-mispricing premium. Although the literature contains ten principal classifications of zombie firms, we report the headline results for three: Banerjee and Hofmann (2020), Favara, Minoiu and Perez-Orive (2021), and Altman, Dai and Wang (2022). The three zombie strategies were chosen to expose the strategy's behavior across three structurally different ICR-family signal constructions; the methodological logic of the selection is set out in Section 2.5. All ten classifications were computed and evaluated as fuzzy scores on two universes (Nikkei 225 and TOPIX) under both equally weighted and capitalization-weighted implementations; the full set is summarized in Section 6.5, and all ten outperform the Nikkei 225 benchmark. The Nikkei 225 equal-weight results form the headline because the Nikkei 225 is the more directly investable benchmark for international allocators and because equal weighting maximizes the influence of the score tails, where the credit-mispricing signal is largest. Across the three classifications, the 26-year point-in-time test delivers a CAGR of 7.2% to 10.2% against 3.8% for the Nikkei 225 benchmark. Sharpe ratios range from 0.41 to 0.52 against 0.28. Carhart (1997) fourfactor regressions on daily excess returns produce annualized alphas of 1.47% to 4.49%, betas below 1.0, value loadings between 0.347 and 0.404 that are statistically significant at the 1% level, and negative and significant momentum loadings. The factor signature is unambiguous: this is a value strategy that buys cheap, beaten-down credit-impaired firms and earns the recovery premium when those firms repair their balance sheets. Because every zombie rule conditions, directly or indirectly, on depressed valuations, Section 6.6 benchmarks the strategies not only against factor models but against investable value placebo portfolios of identical construction; the credit screen adds 2.3 to 3.0 percentage points per annum over these placebos, concentrated in the classifications whose zombies exhibit no elevated exit risk. The remainder of the paper is organized as follows. Section 2 reviews the zombie-firm literature and isolates the five fundamentals that we exploit, with explicit attention to the recovery gap in prior work and to the cycle-robustness argument. Section 3 documents zombie-firm dynamics on the Tokyo Stock Exchange and presents the survival and Cox evidence. Section 4 formalizes the fuzzy scoring methodology. Section 5 describes the data, universe and bias controls and explains the design grid that motivated the headline selection. Section 6 presents the out-of-sample portfolio construction, performance and Carhart attribution.

Section 7 discusses practitioner implications, extensions and limitations. Section 8 concludes.

2. The Zombie-Firm Literature and the Recovery Gap

2.1 The Landscape of Definitions

The zombie-firm literature has produced a sequence of overlapping but distinct definitions. Exhibit 1 summarizes the ten principal classifications and groups them by signal family. Two families dominate: definitions built on the Interest Coverage Ratio (ICR), which ask whether operating earnings cover interest expense, and definitions built on the Interest Rate Gap (IRG), which ask whether the interest the firm actually pays is below the rate it would pay in an arm's-length market. The two families capture conceptually orthogonal dimensions of zombification, which is reflected in the Matthews-correlation evidence reported in Section 3.2. Within the ICR family, the literature has progressively layered market-value, leverage, sales-growth and distress-score filters on top of the basic ICR test. Of the ten classifications shown, the three highlighted are operationalized as fuzzy scores in detail later in the paper. They were selected to span the methodological spectrum of the ICR family — a lean two-factor rule (B&H 2020), a three-factor rule with externally verifiable sales information (Favara), and a hybrid rule with a probability-of-default screen (Altman) — rather than to maximize any single performance metric. Summary statistics for all ten classifications appear in Section 6.5.

Exhibit 1. Principal Zombie Classifications in the Literature by Signal Family

Author(s) and Year	Family	Key criteria
Caballero, Hoshi and Kashyap (2008)	IRG	Negative IRG only — pure subsidized-credit proxy, no viability screen
Fukuda and Nakamura (2011)	IRG	Negative IRG plus EBIT < minimum interest, leverage > 0.5, positive debt growth
Adalet McGowan et al. (2017)	ICR	ICR < 1 for 3 consecutive years AND firm age ≥ 10
Storz et al. (2017)	ICR	DSC < 5% AND negative ROA AND negative net investment, all for 2 years
Banerjee and Hofmann (2018)	ICR	Adalet McGowan criteria plus Tobin's Q below sector median
Banerjee and Hofmann (2020)	ICR	ICR < 1 for 2 years AND Tobin's Q below sector median
Acharya et al. (2020)	IRG	Negative IRG plus 2-year average ICR below sector median plus leverage above median
Favara et al. (2021)	ICR	Leverage above median AND ICR < 1 for 2 years AND negative real sales growth (3 yrs)
Altman et al. (2022) Z / Z″	ICR	3-year MA ICR (EBITDA) < 1 AND 3-year MA Z- or Z″-Score < 1.8
Nakamura (2023)	IRG / ICR	Dual-track: interest-relief criterion OR evergreen-lending criterion

Note. Banerjee and Hofmann (2020), Favara et al. (2021), and Altman et al. (2022) are operationalized as fuzzy scores in Sections 4 through 6. The remaining seven classifications were computed and evaluated as well; headline performance for the full set is reported in Section 6.5. We calculate 11 scores based on 10 strategies as we calculate Altman Z and Z″ scores separately.

2.2 From Macro Diagnostic to Investment Signal

The zombie-firm literature is dominated by a macroeconomic and competition-policy perspective. The consensus diagnosis stresses depressed productivity, capital misallocation, the crowding-out of healthy competitors, and the impairment of Schumpeterian creative destruction (Schumpeter 1942). The Japanese experience since 1989 (collapsing collateral values, deferred non-performing-loan recognition, and politically rational but economically costly forbearance) is the defining case study. This paper takes a different starting point. Where prior work asks what zombie firms do to the macro economy, we ask what they imply for the cross-section of equity returns. The hypothesis is that the same balance-sheet ratios that the literature uses to diagnose distress at the firm level also identify systematic credit mispricing in liquid, exchange-listed equities. Every input is read directly

from the firm's published accounts, and every selection decision is made one stock at a time on the basis of those accounts, with no top-down macro overlay.

2.3 The Recovery and Exit Gap in the Prior Literature

A striking feature of the zombie-firm literature is what it has and has not measured. Considerable energy has been devoted to refining the classification rule, to documenting the count and the sectoral distribution of zombies, and to estimating their macroeconomic externalities. Comparatively little energy has been devoted to the question that matters most for investors: conditional on having been classified as a zombie in year t, what is the firm's empirical probability of defaulting in year t+1, t+2, t+3, and what is its complementary probability of recovering? On what fundamental covariates do those probabilities depend? Two reasons explain the gap. First, most contributions are written from the perspective of a regulator or a macro economist whose object of interest is the aggregate stock of impaired firms, not the conditional trajectory of any individual classified firm. Second, identifying recoveries empirically is hard: one needs a balanced panel that follows classified firms beyond the classification window, with clean accounting data through subsequent quarters and an unambiguous criterion for what counts as recovery. The Tokyo Stock Exchange offers a uniquely long and clean panel on this dimension, with comparable accounting standards and continuous coverage in the Refinitiv / LSEG database from 1995 onwards. Filling this gap is not merely an empirical exercise. The recovery rate of classified zombies is the key economic input into the credit-mispricing thesis. If recovery were rare, the high credit-risk discount embedded in zombie equity prices would be approximately fair, and a long position in elevated-zombie-score firms would earn no positive expected risk-premium gap. If recovery is the modal outcome, the discount is systematically too large, and the cross-section of expected returns contains an accessible premium. Section 3 of this paper documents the latter.

2.4 Cycles of Pronounced Zombification and a Signal That Survives Them

Zombification is cyclical. The post-1989 Japanese experience produced one of the most pronounced and persistent zombification episodes in the modern history of advanced economies. In the years after the collapse of asset prices, depending on the classification used, between 10% and 30% of listed Japanese firms displayed zombie characteristics. Hoshi (2006), Peek and Rosengren (2005), Caballero, Hoshi and Kashyap (2008), Ahearne and Shinada (2005), and Fukuda and Nakamura (2011) document the macro mechanics: depressed bank capital and deferred loan recognition produced sustained evergreen lending, which in turn kept low-productivity firms alive and depressed the equilibrium return on capital across the economy. A second wave appeared in continental Europe after the 2010–12 sovereign-debt crisis (Acharya et al. 2020; Storz et al. 2017), and a third was debated for the post-2009 United States (Favara, Minoiu and Perez-Orive 2021), episodes we cite as context for the classification rules, not as objects of this study. The institutional backdrop is itself in transition: the bank-centered insider ownership system that sustained creditor forbearance has been unwinding since the late 1990s (Franks, Mayer, and Miyajima 2014). Our focus is the Japanese market as it stands today: binary zombie counts on the Nikkei 225 have been low since the mid-2010s, which is precisely the environment in which the continuous score, rather than the

binary flag, carries the investable information. Outside these episodes, the count of classical zombies is low. In the post-pandemic period and in current market conditions, the binary zombie count on the Nikkei 225 is small. A naive binary strategy that buys only flagged firms would therefore be mechanically constrained: too few names, unstable portfolio sizes year by year, and an undefined ranking within the small flagged set. The fuzzy score we develop in Section 4 is unaffected by this constraint. Because it converts the underlying ratios into continuous cardinal scores in the unit interval, every Nikkei 225 constituent receives a well-defined score in every year, and the top decile is always cleanly identifiable. The signal is therefore informative across the cycle: it leverages the rich credit-quality information contained in the underlying ratios whether or not the binary thresholds are crossed. We view this as one of the strongest practical features of the methodology, and one of the main reasons it is suitable for ongoing portfolio management rather than for episodic event studies.

2.5 Two Measurement Traditions and Three Operationalized Classifications

The literature uses two measurement traditions. The Interest Coverage Ratio (ICR) measures how many times operating earnings cover interest expense: ICR = EBIT / Interest Expense. An ICR below 1.0 signals that the firm cannot service its debt out of operations. Adalet McGowan et al. (2017) define a firm as a zombie when its ICR has been below 1.0 for three consecutive years and the firm is at least ten years old, the age filter excluding startups whose growth investments produce temporary losses. Altman et al. (2022) substitute EBITDA for EBIT to neutralize cross-country differences in depreciation accounting. The Interest Rate Gap (IRG), due to Caballero, Hoshi and Kashyap (2008), measures whether the interest the firm actually pays falls below the theoretical minimum it could have obtained from the highest-quality borrower in the market: IRG = Actual Interest Paid − R*. R* is constructed from short-term and long-term prime rates and the minimum convertible-bond coupon over the previous five years, weighted by the firm's short-term loans, long-term loans, and outstanding bonds. A negative IRG indicates subsidized credit, the classic signature of evergreen lending. For investment purposes we exploit the ICR family because its inputs are universally and contemporaneously available for Nikkei 225 constituents through Refinitiv / LSEG. IRG-based information is used as a cross-check on candidate firms. Of the ten classifications in Exhibit 1, the three combinations operationalized below were chosen because, taken together, they exemplify three structurally different ways in which the literature has filtered the ICR signal: by combining it with a market-value test, by combining it with cash-flow and balance-sheet conditions whose evidentiary basis lies outside the firm's accounting cycle, and by pairing it with an explicit probability-of-default model. The remaining seven were computed and evaluated as well; all ten outperform the Nikkei 225 benchmark, and summary statistics for the full set are reported in Section 6.5. The three combinations we operationalize are: Banerjee and Hofmann (2020). Banerjee and Hofmann tighten the Adalet McGowan rule by adding a market-value filter. A firm is a zombie when its ICR has been below 1.0 for two consecutive years and its Tobin's Q is below the sector median. The Q filter is the critical addition for our purposes: a firm trading below the replacement value of its assets is, by definition, undervalued by the equity market. Combining low ICR with low Q therefore identifies firms where the market is simultaneously pricing high credit risk and low fundamental value, the textbook configuration for a contrarian credit-mispricing trade.

Favara, Minoiu and Perez-Orive (2021) require three conditions simultaneously: leverage above the sample median, ICR below 1.0 for two consecutive years, and negative real sales growth over three years. The novelty is the externally verifiable sales-growth criterion. Sales come from end-customers, not from accounting choices or refinancing decisions, and they discriminate cleanly between balance-sheet zombies in the strict sense and young, growing firms whose interest coverage is temporarily depressed by capacity investment. The Favara filter delivers the largest value loading of the three approaches in our backtest: it is the deepest value tilt. Altman, Dai and Wang (2022) demand a three-year moving-average ICR (EBITDA-based) below 1.0 jointly with a three-year moving-average Z-Score (or, in our Japan-adapted variant, Z''-Score) below 1.8. The Z-Score combines liquidity, retained profitability, operating efficiency, leverage and asset turnover; the Z″-Score omits the sales-to-assets term and is therefore more comparable across non-US listings. The Altman dual requirement (a probability-of-default screen and a depth-of-distress screen, both binding simultaneously for three years) is the most demanding in the literature and identifies the smallest, most concentrated set of candidate firms.

This selection logic doubles as a pre-commitment for inference. The credit-mispricing hypothesis predicts a residual over investable value only where an accounting distress signal is combined with a market-value test: a configuration in which the market can misprice fundamentals that do not materialize into exits. The Banerjee and Hofmann (2018, 2020) pair therefore constitutes the hypothesis-bearing set for the spanning tests of Section 6.6; the remaining classifications serve as the comparison field, and all eleven results are reported regardless.

3. Data and Zombie-Firm Dynamics on the Tokyo Stock Exchange

Before any zombie classification is used as an investment signal, three empirical questions need to be answered. First, do different classifications in the literature pick out similar sets of firms, or are they conceptually distinct? Second, is zombie status reversible, and at what frequency? Third, what fundamental covariates drive recovery versus exit, conditional on classification? This section answers all three on the Tokyo Stock Exchange panel for FY1994 through FY2024. The evidence on reversibility and on the drivers of recovery is the empirical bedrock of the credit-mispricing thesis.

3.1 Data: Sources, Universes, and Verification

All data are sourced from the Refinitiv / LSEG database. Three building blocks enter the analysis. The first is market data: daily closing prices for every constituent of the TOPIX and of the Nikkei 225 — including every firm that joined or left either index during the sample — together with the daily levels of both indices themselves. The second is accounting data: annual balance-sheet and income-statement items for the same firms (total assets and liabilities, debt and its maturity structure, EBIT and EBITDA, interest expense, revenue, net income, current assets and liabilities, capital expenditure, depreciation, retained earnings, book equity, and market capitalization), together with the original publication date of each annual report. The third is index composition: the complete joiner-and-leaver history of both indices, from which we reconstruct the exact membership at every rebalancing date; the reconstructed Nikkei 225 panel contains exactly 225 names at every year-end.

The two universes play distinct roles throughout the paper. The broader TOPIX panel — 1,200 to 2,200 classifiable non-financial firms per year, against roughly 200 on the Nikkei 225, with binary zombie incidence below 5% — provides the statistical power for the evidence in this section: concordance, reversibility, survival, and the drivers of recovery. The Nikkei 225 is the investable universe on which the portfolio strategy of Sections 5 and 6 is built and tested. Every Nikkei constituent is a TOPIX constituent and the classification rules are identical on both panels, so the evidence measured here transfers directly to the portfolio universe. Section 7.2 adds a pricing counterpart to this division of labor: the part of the zombie premium that survives value spanning is confined to the blue-chip segment, so the restriction of the strategy to the Nikkei 225 is substantive rather than merely practical.

The complete data set (membership reconstruction, accounting panels, price histories, and publication dates) was verified end-to-end by our data scientist, including a reconciliation of the vendor's index event feeds against reported index counts and an audit of the announcement-date fields on which the availability rule of Section 5.1 operates.

3.2 Concordance Across Classifications

The Matthews Correlation Coefficient (MCC; Matthews 1975) is the standard pairwise concordance metric for binary classifications and is robust to the strong class imbalance that characterizes binary zombie classifications: fewer than 5% of TOPIX constituents in most years. Exhibit 2 reports the pairwise MCC matrix across the ten classifications. Within-family correlations are high (BH18/BH20 = 0.73; BH18/McGowan = 0.68), while cross-family correlations between IRG-based and ICR-based rules are essentially zero. The three ICR-based rules we operationalize in Section 6 are conceptually related but empirically distinct enough to deliver complementary information. A principal component analysis of the eleven fuzzy scores summarizes this structure in two numbers. The first component explains 55 percent of the cross-sectional variance on average and loads positively on all eleven scores in every year: a common distress factor. The second component explains 19 percent, and its loadings separate exactly the two measurement traditions — positive on the interest-rate-gap rules, negative on the coverage-based block, with the hybrid Acharya rule in between; the eigenvalue-ratio criterion of Ahn and Horenstein (2013) selects one to two factors. The loadings appear in Panel B of Table A.4 (Factor Structure of the Eleven Fuzzy Scores) in Appendix A.4. The two-family taxonomy of Section 2.5 is thus a statistical property of the data, not a classification choice.

Exhibit 2. Pairwise Matthews Correlation Coefficients Across Ten Binary Zombie Classifications, TOPIX Panel

	CAB	F-N	STZ	MCG	BH18	BH20	ACH	FAV	A-Z	A-Z″
F-N	0.12									
STZ	0.00	0.03								
MCG	0.00	0.01	0.23							
BH18	0.00	0.02	0.15	0.68						
BH20	0.00	0.01	0.18	0.49	0.73					
ACH	0.35	0.36	0.01	0.01	0.00	0.01				
FAV	0.00	0.00	0.26	0.52	0.31	0.42	0.04			
A-Z	0.00	0.01	0.21	0.41	0.32	0.30	0.02	0.35		
A-Z″	0.00	0.01	0.18	0.48	0.25	0.22	0.03	0.43	0.68	
NAK	0.33	0.04	0.09	0.26	0.18	0.26	0.11	0.27	0.18	0.22

Note. The Altman (2022) Z and Z″ variants are shown as separate columns. Codes: CAB = Caballero et al. (2008), F-N = Fukuda and Nakamura (2011), STZ = Storz et al. (2017), MCG = Adalet McGowan et al. (2017), BH18 = Banerjee and Hofmann (2018), BH20 = Banerjee and Hofmann (2020), ACH = Acharya et al. (2020), FAV = Favara et al. (2021), A-Z = Altman Z-Score (2022), A-Z″ = Altman Z″-Score (2022), NAK = Nakamura (2023). Availability-masked panel, FY1995–FY2023. Source: Refinitiv / LSEG; authors' calculations.

3.3 Reversibility: Recovery Is the Modal Outcome

The most important empirical fact for the credit-mispricing thesis is that zombie status, on the Tokyo Stock Exchange, is typically short-lived. Across the ten classifications, the median duration of an unbroken zombie spell is one to two years; the longer two-year median spells occur under Adalet McGowan and the two Altman variants. Annual exit intensities from the zombie pool are low throughout: at most 1.7% (Storz et al.), typically below 1.2%, and zero in the sparsely populated interest-rate-gap pools, against roughly 0.6% per annum for never-classified firms. Annual recovery rates from the zombie state range from 33% to 73%, and annual relapse probabilities range from approximately 1% to 5.5%; Table A.7 (Survival and Recovery by Classification) tabulates these statistics for every classification. These numbers leave the modal outcome unambiguous: in any given year, a firm classified as a zombie is far more likely to remain operational (and, in most cases, to subsequently recover) than to disappear through bankruptcy or delisting. Market exit in this panel is delisting of any cause and therefore includes takeovers; separating failure-type exits from acquisitions requires delisting-reason data beyond the present panel. The direction of the resulting bias is conservative for our thesis: counting takeovers as exits raises measured zombie exit intensity, so the finding that classified companies rarely exit would, if anything, strengthen under the separation.

The implication for the equity market is direct. The fundamental profile that earns a firm a zombie label (low interest coverage, depressed valuation, elevated leverage) is exactly what a value investor screens for and it is also the profile to which equity investors apply substantial credit-risk discounts. If the discount were set equal to the probability-weighted expected loss from default, it would be small, because default is rare. The discount instead appears to be set as if default were the central scenario, when it is in fact the tail outcome. The gap between this perceived default risk and the realized default risk is the credit-mispricing premium.

This is also where our methodology directly addresses the recovery gap in the prior literature mentioned in Section 2.3. The earlier literature has tended to take the count of zombies as its principal

outcome of interest. The cross-section of expected equity returns, however, depends not on how many firms are zombies in year t but on what happens to those firms in years t+1, t+2 and beyond. The recovery numbers we report make the case that, for a long-only equity investor, the relevant statistic is not the prevalence of zombification but its reversibility — and reversibility on the Tokyo Stock Exchange is high.

3.4 Kaplan–Meier Survival and the Heterogeneity of Distress

The Kaplan-Meier estimator provides a non-parametric, right-censored view of cumulative survival from the date of first zombie classification. Comparing zombie and never-zombie cohorts on the TOPIX panel from 1995 through 2025 yields a striking heterogeneity across classifications; Panel A of Table A.7 (Survival and Recovery by Classification) reports the statistics. At one end, the flow-based Storz et al. rule selects firms with materially elevated exit risk: its zombie group trails the never-zombie group by 6.9 percentage points in five-year survival and by 10.5 points at ten years, and its annual exit intensity is 2.9 times the never-zombie rate. The Altman variants produce intermediate gaps (3.5 to 4.2 points at ten years), Favara et al. a smaller one (2.4 points). At the other end sit the interest-coverage-and valuation rules: the Banerjee and Hofmann (2020) portfolio exits at the same annual rate as never-classified firms (intensity ratio 1.00) and trails by only 1.7 points in ten-year survival, while the Banerjee and Hofmann (2018) cohort exits at half the never-zombie rate. From an investment perspective, a narrow survival gap is not a weakness of the classification: it is precisely where credit mispricing is most likely. Firms classified as Banerjee and Hofmann (2020) zombies look distressed by accounting standards but exit no more often than the average TOPIX constituent. The market discount that their accounting profile attracts is therefore largely unwarranted, and that is exactly the population from which our top-decile portfolio draws.

3.5 Cox Proportional Hazards: What Drives Recovery?

The Cox proportional-hazards model identifies the firm-level covariates that drive the conditional probability of market exit from the zombie pool, controlling for time. To avoid mechanical circularity, we use covariates that are deliberately absent from the classification formulas themselves: asset productivity (proxied by asset turnover, as employee counts are not in the panel), liquidity (proxied by the current ratio, as cash and equivalents are not separately available), capital intensity (non-current over total assets), leverage (total debt over book equity), and within-sector revenue share as a proxy for market position. Covariates are winsorized at the 1st and 99th percentiles, transformed by the signed logarithm where skewed, and z-standardized within year; one pooled proportional-hazards model is fitted per classification in the zombie cohort, for the six classifications with at least fifty observed exits, and hazard ratios are reported per one standard deviation of the covariate. Panel B of Table A.7 (Survival and Recovery by Classification) reports the estimates where two findings dominate.

Capital intensity is the hidden lifeline. Asset-heavy zombies exit less frequently in every specification: hazard ratios lie between 0.72 and 0.87. The mechanism is the classical collateral channel: high liquidation costs and pledgeable collateral give creditors an incentive to maintain forbearance rather than recover their claims through bankruptcy.

Liquidity is the second protective covariate, in five of the six classifications (hazard ratios 0.69 to 0.83; the Banerjee-Hofmann 2020 pool is the exception), completing the classical forbearance configuration in which pledgeable collateral and adequate liquidity give creditors both the means and the incentive to wait. The remaining covariates are close to flat: leverage carries hazard ratios of 0.86 to 1.09, sector market share 0.86 to 1.07, and asset productivity is mixed (0.84 to 1.15). Because the entire zombie population is highly levered by construction, no single balance-sheet dimension is sufficient to discriminate; the score must combine multiple dimensions, which is what the fuzzy aggregation in Section 4 delivers.

4. Fuzzy Scoring: From Binary Flag to Investable Signal

4.1 Why the Binary Flag Is Unsuitable for Portfolio Construction

Every classification in the literature reviewed in Section 2 produces a binary output: a firm is or is not a zombie at a given date. This is appropriate for the macro question (what fraction of the corporate sector is impaired) but it is inadequate for portfolio construction, for two reasons. First, the binary flag discards severity information. Consider two firms whose ICR has been below 1.0 for two consecutive years. Firm A has ICR = 0.10; Firm B has ICR = 0.95. Both receive the same flag, yet Firm A is severely distressed and Firm B is marginal. Treating them identically eliminates the cross-sectional dispersion that is the very source of the investment signal. Second, the binary flag does not produce a ranking. A long-only portfolio strategy needs to choose a fixed number of names from a defined universe; it cannot operate on a population that is split arbitrarily into flagged and unflagged buckets whose sizes vary year by year. Without a continuous score, the universe of investable names is unstable, the portfolio is uninvestable, and the strategy is untestable. The fuzzy score removes both limitations in a single, mathematically transparent transformation.

4.2 The Five Fundamental Inputs

From an investment standpoint, the literature converges on five core balance-sheet inputs: the interest coverage ratio, Tobin's Q, leverage, real sales growth, and the Altman Z-Score (or Z″-Score). The case for using these five (and not a wider set of accounting ratios or technical indicators) rests on three observations.

First, every ratio is a direct expression of a credit-risk fundamental. ICR captures debt-service capacity, leverage captures balance-sheet vulnerability, the Z-Score (Altman 1968) consolidates the probability of default, sales growth captures going-concern viability, and Tobin's Q (Tobin 1969) captures the market's joint pricing of those fundamentals. Taken together, these five inputs correspond directly to the credit-quality variables emphasized in the classical security-analysis tradition (Graham and Dodd 1934).

Second, all five inputs are available on identical terms for every firm in the sample. They come from audited annual statements, are defined homogeneously across Nikkei 225 constituents, and are covered by the data described in Section 3.1 for the full sample period — including for firms that later left the index, which is what keeps the backtest free of survivorship bias at the data level.

The importance of this discipline for long-horizon return evidence is well established (Brown, Goetzmann and Ross 1995; Goetzmann and Jorion 1999).

Third, these ratios predict survival, not only returns. The Cox proportional-hazards evidence of Section 3.5 shows that balance-sheet structure (capital intensity and liquidity in particular) significantly predicts whether a classified firm exits or recovers, even after conditioning on the classification rule itself. The five inputs therefore carry exactly the information a credit-mispricing strategy needs.

4.3 The Four-Stage Transformation

The fuzzy score is built in four steps in each calendar year, using only information available at the rebalancing date. The intuition of each step is given first, followed by the formal definition. Stage 1: Bound the extreme outliers. In any cross-section a small number of firms have accounting ratios so extreme that they would dominate any aggregation. We therefore winsorize each of the five raw inputs at its 1st and 99th percentile within the year. No firm is dropped from the sample; only its inputs are capped at the corresponding percentile values. Stage 2: Put everything on the same scale. The five inputs are measured in different units: a yen of EBIT is not a yen of total assets, and a percentage growth rate is not a ratio. We therefore apply a signed-log transformation x -> sign(x) log(1 + |x|) to compress fat tails where the annual cross-sectional skewness exceeds 0.5 in absolute value, and then z-standardize within the year. After this step the cross-sectional mean of each variable is zero and the cross-sectional standard deviation is one, so the five inputs can be combined on equal footing. Stage 3: Translate into a 0-to-1 score. We map each standardized value z to a number between 0 and 1 using the logistic sigmoid f(z) = 1 / (1 + exp(+/- z)). The sign in the exponent is chosen so that more zombie-like values (low ICR, low Tobin's Q, high leverage, weak sales growth, depressed Z-Score) always map to higher scores. The sigmoid has its sharpest gradient around the median firm, exactly where the cross-sectional ranking is hardest to call and where extra resolution is most valuable. Stage 4: Combine sub-scores into one composite. When a classification has more than one criterion, we combine the sub-scores into a composite by their geometric mean. The geometric mean has the useful property that weakness on any one dimension drags the composite down: a firm that scores well on three dimensions but very poorly on one ends up with a low composite. This matches the structure of the multi-criterion zombie definitions, which require every condition to be met simultaneously, not on average.

4.4 Sub-Score Specifications by Classification

The three classifications we operationalize differ in their input vectors. Banerjee and Hofmann (2020) take the geometric mean of the ICR sub-score and the Tobin's Q sub-score. Favara et al. (2021) take the geometric mean of the ICR sub-score, the leverage sub-score, and the real sales-growth sub-score. Altman et al. (2022) combine the three-year EBITDA-based ICR sub-score with the three-year Z-Score sub-score; the Z″-Score variant combines the ICR sub-score with the Z″-Score sub-score instead. In every case the inputs are the same balance-sheet quantities used by the

binary rule; only the mapping from observation to score has changed.[1] One rule in the wider set departs from this pattern: Nakamura (2023) classifies through two alternative routes, subsidized credit (a negative interest-rate gap) or balance-sheet distress (uncovered interest combined with rising debt), and a firm qualifies if either route fires. The fuzzy translation preserves the disjunction: the Nakamura score is the maximum of the two route scores — the interest-rate-gap sub-score on the one hand, the geometric mean of the ICR and debt-growth sub-scores on the other. The maximum operator is the fuzzy counterpart of a logical OR, exactly as the geometric mean acts as a soft counterpart of AND. Exhibits 1 and 3 thus describe different objects: Exhibit 1 lists the binary original rules with their thresholds and multi-year conditions, while Exhibit 3 shows how the four-stage transformation maps each rule's cardinal inputs into sub-scores.

Exhibit 3. Sub-Score Composition by Classification

Classification	Sub-score composition
Banerjee and Hofmann (2020)	ICR sub-score × Tobin's Q sub-score
Favara et al. (2021)	ICR sub-score × Leverage sub-score × Real Sales Growth sub-score
Altman et al. (2022) Z-Score	ICR (EBITDA) sub-score × Z-Score sub-score
Altman et al. (2022) Z″-Score	ICR (EBITDA) sub-score × Z″-Score sub-score
Nakamura (2023)	max(IRG sub-score, ICR sub-score × Debt-Growth sub-score)

Note. All sub-scores are constructed via the four-stage transformation of Section 4.3. × denotes the geometric mean. Nakamura's rule is the one disjunction in the set; its score takes the maximum of its two route scores, the fuzzy counterpart of a logical OR. The remaining classifications combine their sub-scores by geometric mean throughout.

4.5 Using Fuzzy Membership for Portfolio Formation

Our contribution is not a new statistical technique but a new use of the literature's own measurements. Prior empirical work treats the classifications as descriptive: how many zombie firms exist, in which sectors, with what macro consequences. For portfolio formation, the binary rules must instead be converted into a ranking, and the conversion should add as little discretion as possible. The four-stage transformation of Section 4.3 is built on the observation that the underlying ratios are themselves cardinal measures of fundamental credit quality and to extract that cardinal information through a transformation that preserves the structure of the binary rules. The resulting signal is ordinal (every constituent is rankable), interpretable (each sub-score traces to an audited line item, no machine-learning black box), and replicable. Equally important, the fuzzy construction generates the cycle-robustness property highlighted in the introduction: in years in which the binary count of zombies is high, the score concentrates the top decile in obviously distressed firms; in years in which the count is low (such as the current Japanese environment) the score still identifies the most credit-impaired firms within the cross-section, allowing the strategy to be operated continuously rather than episodically. The transformation as a whole, not any single design choice within it, carries the results. As a direct check, we replaced the two aggregation stages (the logistic map of Stage 3 and the geometric mean of Stage 4) with the first principal component of each classification's sub-scores, estimated on the pooled five-year formation panel in the spirit of the statistical factor models of Connor and Korajczyk (1986). The rankings barely move: Spearman

[1] The binary originals of Banerjee and Hofmann (2018, 2020) and Acharya et al. (2020) define Tobin's Q, ICR, and leverage relative to sector medians; our operationalization uses the cardinal quantities and delegates all relativization to the within-year standardization of Stage 2. Recomputing all eleven variants under the sector-relative reading (holding everything else fixed) changes no CAGR by more than 0.7 percentage points; the headline conclusions are invariant to this choice.

correlations between the fuzzy and the principal-component formation rankings reach 0.99 to 1.00 for the two-input classifications and 0.78 to 0.97 for the multi-input rules, every principal-component portfolio beats the benchmark, and compound annual returns shift by at most 1.7 percentage points in either direction. Panel A of Table A.4 (Fuzzy Aggregation versus First Principal Component) in Appendix A.4 reports the comparison for all ten classifications. We retain the fuzzy aggregation for three reasons. Principal components maximize explained variance, not distress relevance (Lettau and Pelger 2020). Their loadings are estimated from the data and can rotate across samples, whereas the fuzzy weights are fixed and identical in every year and market. And the one intrinsically nonlinear rule in the literature — Nakamura's fuzzy-OR aggregation (Exhibit 3, Sub-Score Composition by Classification) — has no linear counterpart.

4.6 Relation to Fuzzy Zombie Identification in the Prior Literature

The idea of relaxing binary zombie classifications into continuous ones is not without precedent. Caballero, Hoshi and Kashyap (2008) already treat the set of subsidized-credit firms as a fuzzy set on their one-dimensional interest-rate-gap criterion, and Acharya, Eisert, Eufinger and Hirsch (2019) adopt the same device. The most systematic treatment is Mingarelli, Ravanetti, Shakir and Wendelborn (2022), who formalize arbitrary crisp zombie identifications as indicator functions on a vector of firm characteristics and generalize them to fuzzy identifications by replacing the indicator with a membership function mapping onto the unit interval, aggregated (as in our Stage 4) by the geometric mean of the univariate kernels. Their motivation, however, is diagnostic rather than investive: the fuzzy extension is introduced to increase the consistency of zombie shares measured under different definitions on euro-area firm data, and every design choice follows from that objective. Three of those choices make the Mingarelli et al. (2022) construction unsuitable as a portfolio signal. Mingarelli et al. (2022) employ a piecewise linear membership function anchored at the original binary threshold and truncated at an upper threshold (typically the cross-sectional median) so that the binary zombie count is preserved by construction and roughly half of the cross-section receives a score of exactly zero. The choice is deliberate: it adds close-to-zombie firms to a diagnostic count without disturbing that count. But it leaves the healthier half of the universe unranked and creates a mass point at zero, which is fatal for decile selection in years with few classical zombies. Our sigmoid mapping of within-year standardized inputs assigns every constituent a strictly interior score, with maximal resolution around the cross-sectional median, precisely where a decile boundary must discriminate. The Mingarelli et al. (2022) kernel is anchored in absolute thresholds, and the authors themselves document the consequence: their fuzzy Caballero, Hoshi and Kashyap (2008) score drifts as the general level of interest rates changes, because a fixed upper threshold mechanically reclassifies firms when the underlying distribution compresses. Our procedure standardizes within the year, so the score is a relative position in the annual cross-section by construction. The two normalizations answer different questions. An absolute score measures how zombified the economy is (the diagnostic question). A relative score ranks firms against their contemporaries (the investment question), and it is the source of the cycle-robustness property documented in Sections 4.5 and 6.4. Mingarelli et al. (2022) validate the fuzzy extension by its internal consistency: the overlap between competing zombie definitions rises as the score threshold is lowered. In their tran-

sition-matrix analysis they further observe that firms with lower fuzzy scores recover more frequently, and flag these transition dynamics as future research. This paper operationalizes exactly that observation. The survival and hazard evidence of Section 3 quantifies the recovery asymmetry on a three-decade Tokyo Stock Exchange panel, and the backtest of Section 6 verifies out of sample that the asymmetry is accessible. In this sense the contribution is complementary rather than competing: Mingarelli et al. (2022) establish that fuzzy scores are the conceptually appropriate representation of an inherently graded phenomenon; we establish that, once redesigned for ranking rather than counting, they carry investable information.

5. The Point-in-Time Investment Protocol

5.1 Timing, Availability, and Bias Controls

Throughout, we label Japanese fiscal years by their April start in the local convention: fiscal year t (FYt) begins on 1 April of calendar year t and ends on 31 March of calendar year t+1, with financial statements published by the summer of calendar year t+1; portfolio years are calendar years. This April-start convention covers the large majority of the index: roughly 85 percent of Nikkei 225 constituents close their fiscal year in March, while most of the remainder — about 15 percent in recent years, fewer earlier in the sample — report on a calendar-year basis. The availability rule below treats both groups identically because it operates on each firm's actual publication date rather than on a common fiscal calendar: a December statement published in the spring becomes usable at the same year-end as a March statement published in the summer, and look-ahead is excluded in both cases. The sample covers FY1994 through FY2024 (statements ending March 1995 through March 2025). Financial services companies are excluded from the sample throughout the paper because their balance-sheet structure makes leverage and ICR mechanically non-comparable to non-financial firms. The investable universe at every rebalancing date is the Nikkei 225 as it was actually constituted on that date, including all index additions and deletions over the sample period. Firms that were added to the index after 1995 enter the universe only at their inclusion date; firms that were deleted leave the universe at their deletion date and contribute their realized return up to that point. This composition treatment eliminates two well-known biases by construction. Survivorship bias is eliminated because failed or delisted firms appear in the universe of their tenure years and contribute their realized returns; they are not retroactively dropped. Look-ahead bias is eliminated because the fuzzy score in any year t is computed exclusively from financial-statement information that was publicly available at the rebalancing date in year t: each annual statement enters the calculations only from its original balance-sheet announcement date (where the announcement date is missing, the statement is treated as usable only from the following year-end, a conservative fallback). No future information enters the ranking or the formation-period averaging.

5.2 The Five-Year Formation Period and Annual Rebalancing

The investment protocol is a strict, mechanical, no-optimization procedure: no weight, threshold or window is fitted to returns, the headline classifications are selected on methodological grounds (Section 2.5), and the sensitivity of results to the formation window and the portfolio cut-off is reported in full in Appendix A. We therefore describe the exercise as a point-in-time backtest. Every

input is restricted to information publicly available at the formation date. We use "out of sample" only in the relative sense that each portfolio year lies outside the formation window from which its ranking was constructed.

In each year from FY1994 onwards, every Nikkei 225 constituent at year-end is ranked by its annual fuzzy score, computed as described in Section 4 from the most recent statement available under the availability rule. In each calendar year y from 2000 onwards, the formation-period score is the simple arithmetic average of the firm's annual fuzzy scores over the preceding five years (y-5 through y-1). Firms with fewer than five years of available scores are excluded from the ranking for that year. On the basis of the five-year-average score, the 23 highest-scoring firms among those with complete data for all five formation years (approximately the top decile of the index) are selected as the portfolio for the forthcoming year y. The portfolio is equally weighted at inception. The portfolio is held without intervention from 1 January of year y until 31 December of year y. At year-end the portfolio is rebalanced: the new five-year-average ranking is computed using data through year y, and the new top decile is constructed for year y+1.

Transaction costs of 50 basis points are charged on the value of every purchase and every sale at every rebalancing date. Delistings and index deletions are handled as follows: a constituent delisted during the holding year remains in the portfolio at its last traded price until the end of the portfolio calendar year. The position is carried at that frozen value, with no intra-year reinvestment and its proceeds re-enter the regular year-end rebalance, at which the standard transaction-cost schedule applies; a firm deleted from the index but still listed is held to the regular year-end rebalance. Reinvesting delisting proceeds pro rata across the surviving holdings at the delisting date instead changes headline CAGRs by fewer than 20 basis points per annum (7 to 19 bp across the three headline classifications). The 50 bp charge is generous relative to actual implementation costs in liquid Nikkei 225 names; it is set deliberately high so that reported alpha is not the artifact of an unrealistic frictionless assumption. Returns are computed throughout from Refinitiv adjusted closing prices in local currency, and the benchmark is the corresponding price index; dividends are thus excluded symmetrically on both sides of every comparison — including the spanning tests of Section 6.6, where strategy and placebo portfolios share the same universe and construction.

Because the formation period comprises FY1994 through FY1998 (statements ending March 1995 through March 1999, all published before 1 January 2000) the first investable portfolio is held in calendar year 2000. Using FY1999 (ending March 2000, published in the summer of 2000) would violate the availability rule; it enters only the portfolio held in 2001. The strategy is operated annually through 2025, producing a continuous 26-year out-of-sample track record. Every year in that track record is out of sample in the strict sense relative to the formation window from which its portfolio was constructed.

6. Investment Results

This section reports the results in a fixed order, under one baseline specification throughout: the three headline classifications of Section 2.5, the five-year formation average, the 23 highest-scoring names, equal weighting, annual rebalancing, and 50 basis points of transaction costs on every trade.

Sections 6.1 through 6.4 present annual returns, risk-adjusted performance, the Carhart factor attribution, and behavior across market regimes for the three headline classifications. Section 6.5 widens the evaluation to all ten classifications in the literature under the identical protocol, and Section 6.6 asks the decisive economic question: whether the strategies deliver anything beyond investable value. Sensitivity to the formation window and to the portfolio cut-off is reported in Appendix A.1 and A.2; the capitalization-weighted and TOPIX-based implementations are summarized in Section 7.2.

6.1 Annual Returns, 2000–2025

Exhibit 4 presents the simple annual returns of the three fuzzy-score strategies alongside the Nikkei 225 benchmark, all in Japanese yen and all net of the 50 bp transaction-cost charge. All three strategies deliver an arithmetic mean return roughly twice that of the benchmark over the 26-year out-of-sample window. The Banerjee and Hofmann (2020) strategy delivers an arithmetic mean of 14.0% (CAGR 10.2%), the Altman et al. (2022) ZScore strategy 12.0% (CAGR 8.9%), and the Favara et al. (2021) strategy 10.7% (CAGR 7.2%), against 6.2% and 3.8% respectively for the Nikkei 225 benchmark.

Exhibit 4. Annual Simple Returns, 2000–2025 (JPY)

Year	Nikkei 225	Banerjee and Hofmann (2020)	Favara et al. (2021)	Altman et al. (2022) Z-Score
2000	-27.2%	10.0%	-1.2%	19.7%
2001	-23.5%	-6.9%	-18.4%	-6.3%
2002	-18.6%	-15.2%	-24.3%	-13.6%
2003	24.5%	88.4%	100.3%	71.3%
2004	7.6%	19.7%	16.8%	15.8%
2005	40.2%	91.5%	76.6%	100.8%
2006	6.9%	-12.1%	-15.1%	-14.9%
2007	-11.1%	-16.3%	-10.2%	-4.8%
2008	-42.1%	-37.4%	-30.3%	-31.7%
2009	19.0%	5.7%	0.3%	-1.8%
2010	-3.0%	6.3%	1.0%	0.7%
2011	-17.3%	-13.6%	-14.1%	-15.6%
2012	22.9%	22.0%	15.1%	23.8%
2013	56.7%	67.0%	72.0%	56.4%
2014	7.1%	18.2%	11.0%	9.5%
2015	9.1%	8.1%	9.5%	7.8%
2016	0.4%	2.3%	7.5%	-1.9%
2017	19.1%	20.6%	15.8%	9.2%
2018	-12.1%	-23.4%	-16.7%	-10.4%
2019	18.2%	4.1%	4.9%	5.7%
2020	16.0%	-10.5%	-12.6%	-11.4%
2021	4.9%	37.1%	25.5%	32.2%
2022	-9.4%	10.2%	7.9%	11.5%
2023	28.2%	36.0%	28.8%	33.9%
2024	19.2%	25.9%	9.9%	2.7%
2025	26.2%	26.4%	18.6%	23.5%
Arith. mean	6.2%	14.0%	10.7%	12.0%
CAGR	3.8%	10.2%	7.2%	8.9%

Note. Strategies are equally weighted, top-decile, annually rebalanced, net of a 50 bp transaction-cost charge. Benchmark: Nikkei 225. Source: Refinitiv / LSEG; authors' calculations.

6.2 Risk-Adjusted Performance

Annual Sharpe ratios, computed on daily excess returns over the one-year JPY government yield, are presented in

Exhibit 5. Annual Sharpe Ratios on Daily Excess Returns, 2000–2025

Year	Nikkei 225	Banerjee and Hofmann (2020)	Favara et al. (2021)	Altman et al. (2022) Z-Score
2000	-1.32	0.47	0.12	0.80
2001	-0.80	-0.12	-0.59	-0.13
2002	-0.69	-0.47	-0.76	-0.47
2003	1.09	2.55	2.63	2.45
2004	0.51	0.95	0.78	0.85
2005	2.63	3.64	3.27	3.98
2006	0.44	-0.40	-0.57	-0.57
2007	-0.60	-0.63	-0.42	-0.16
2008	-1.00	-0.79	-0.73	-0.80
2009	0.77	0.33	0.12	0.04
2010	-0.05	0.39	0.14	0.13
2011	-0.72	-0.39	-0.49	-0.59
2012	1.36	1.06	0.89	1.39
2013	1.85	1.93	2.08	1.70
2014	0.45	0.90	0.63	0.56
2015	0.53	0.50	0.58	0.49
2016	0.15	0.23	0.41	0.06
2017	1.58	1.62	1.25	0.82
2018	-0.58	-1.31	-0.93	-0.52
2019	1.34	0.34	0.41	0.50
2020	0.73	-0.22	-0.32	-0.25
2021	0.36	1.47	1.15	1.38
2022	-0.39	0.62	0.54	0.72
2023	1.67	1.93	1.77	2.02
2024	0.82	1.04	0.54	0.24
2025	1.11	1.24	1.07	1.35
Full period	0.28	0.52	0.41	0.49

Note. Risk-free rate: one-year JPY government bond yield. Source: Refinitiv / LSEG; authors' calculations.

Over the full 26-year sample the strategies' Sharpe ratios range from 0.41 to 0.52, against 0.28 for the Nikkei 225. The improvement is consistent across sub-periods: the strategies dominate in expansionary regimes (2003–2005, 2012–2013, 2021, 2023) and exhibit drawdowns broadly comparable to the benchmark in crisis years (2008, 2011, 2018, 2020). In 2008 the strategies in fact fell somewhat less than the benchmark; deep-value portfolios nevertheless remain the most exposed to liquidity-driven sell-offs, and the crisis-year numbers should be read with that structural caveat in mind.

6.3 Carhart Four-Factor Attribution

The raw and risk-adjusted results establish that the strategies outperform the benchmark on both an absolute and a risk-adjusted basis, but they do not establish that the outperformance is genuine alpha rather than compensation for systematic factor exposure. We address this by regressing daily strategy excess returns on the Fama and French (1993) and Carhart (1997) four factors: market, size (the small-firm effect), value (the value effect), and momentum (the momentum effect), using

Newey and West (1987) heteroscedasticity- and autocorrelation-consistent standard errors. The factor data are the daily Japanese Fama-French three factors and the daily Japanese momentum factor sourced from Kenneth R. French's data library (French n.d.). Results appear in Exhibit 6.

Exhibit 6. Carhart (1997) Four-Factor Regression on Daily Excess Returns

Factor	Banerjee and Hofmann (2020)	Favara et al. (2021)	Altman et al. (2022) Z-Score
Alpha (annualized)	4.49%	1.47%	3.50%
Newey–West t-stat	1.429	0.494	1.251
Beta	0.959 ***	0.895 ***	0.865 ***
t-stat	47.05	38.86	42.07
Size	0.247 ***	0.289 ***	0.214 ***
t-stat	5.24	6.05	5.24
Value	0.404 ***	0.391 ***	0.347 ***
t-stat	11.48	11.68	10.70
Momentum	−0.135 ***	−0.110 **	−0.073 *
t-stat	−2.89	−2.54	−1.87
R-squared	0.593	0.567	0.585
Observations	6,373	6,373	6,373

Note. Beta = market factor; Size = small-firm effect (SMB); Value = value effect (HML); Momentum = momentum effect (MOM). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Newey–West standard errors use five lags; results are insensitive to lag choices between three and ten (t-statistics move by less than 0.04 across that range). Source: Refinitiv / LSEG; authors' calculations. Portfolio returns are price-based while the factor series are total-return-based; the annualized intercepts are therefore conservative, understated by roughly the portfolio dividend yield.

Three observations dominate the factor attribution. Annualized Alpha is positive across all three classifications, with a maximum of 4.49% per annum for Banerjee and Hofmann (Newey-West t = 1.43; economically meaningful but short of statistical significance at conventional levels). The value loading is large, positive, and statistically significant at the 1% level for every strategy, ranging from 0.347 (Altman) to 0.404 (Banerjee and Hofmann). This is the empirical fingerprint of the credit-mispricing thesis: the strategy systematically buys firms that are "cheap" on fundamental valuation, which is exactly what the value effect rewards. The momentum loading is negative and statistically significant for every strategy (−0.135 to −0.073). This too is consistent with the credit-mispricing thesis: the strategy buys firms whose stock prices have been beaten down, so it loads negatively on the recent-winner factor. A contrarian, fundamental, deep-value strategy could not look any other way. The market beta sits below 1.0 for all three strategies (0.865 to 0.959), so the absolute outperformance is achieved with slightly less market exposure than the benchmark. On the corrected panel the size loading is positive and significant for all three strategies: part of the raw outperformance reflects smaller-cap exposure within the index. The R-squared of approximately 0.57 to 0.59 indicates that a meaningful share of daily variance is explained by the four factors, but the alpha space remains as a non-trivial residual.

6.4 Performance Through the Cycle

Inspection of Exhibits 4 and 5 shows that the full-sample numbers are not driven by a handful of exceptional years. The three strategies outperform the benchmark in 16 to 18 of the 26 sample years. They participate in the 2003 and 2005 rebounds with returns of 71% to 101%, in the 2013 Abenomics rebound with returns of 56% to 72%, and in the 2021 reflation with returns of 26% to 37%. They underperform in years that are systematically valueunfriendly: 2006, 2009, 2019, 2020. The cycle-robustness argument from Section 4.5 finds its most direct empirical expression here. In years

in which the binary count of Banerjee and Hofmann (2020) zombies on the index is high (early 2000s, post-2008), the strategy's top decile is dominated by obviously distressed names and earns large recovery premiums. In years in which the binary count is low (post-2013 for some classifications; almost all years for Altman et al. 2022 given its stringent conjunctive criteria), the fuzzy score still produces a complete cross-sectional ranking, and the top decile is still investable. The strategy is therefore operable across the cycle, without recourse to ad-hoc rules about what to do in low-zombie-count years.

6.5 Robustness Across the Literature

The preceding subsections document the headline results for the three classifications selected on methodological grounds in Section 2.5. This subsection widens the lens: we computed and evaluated fuzzy-score strategies for every classification in the literature under the identical protocol — the same universe, formation window, weighting, and cost assumptions. Exhibit 7 reports the headline performance metrics for the full set (ten classifications, shown in eleven variants because the Altman rule is reported in both calibrations) at the paper baseline (five-year formation period, top decile, equal weighting, Nikkei 225, 20002025, net of 50 basis points transaction costs).

Exhibit 7. All Ten Zombie Classifications — Headline Performance

Classification	Family	CAGR	Sharpe
Nikkei 225 (Benchmark)	–	3.8%	0.28
Caballero et al. (2008)	IRG	5.6%	0.36
Fukuda and Nakamura (2011)	IRG	5.8%	0.36
Storz et al. (2017)	ICR	5.1%	0.33
Adalet McGowan et al. (2017)	ICR	8.9%	0.47
Banerjee and Hofmann (2018)	ICR	10.8%	0.54
Banerjee and Hofmann (2020) *	ICR	10.2%	0.52
Acharya et al. (2020)	IRG	7.9%	0.44
Favara et al. (2021) *	ICR	7.2%	0.41
Altman et al. (2022) Z-Score *	ICR	8.9%	0.49
Altman et al. (2022) Z″-Score	ICR	7.5%	0.43
Nakamura (2023)	IRG	9.1%	0.48

Note. Long-only top-decile portfolio of the fuzzy score, rebalanced annually on the Nikkei 225, 2000–2025, equal weighting, net of 50 bp transaction costs. An asterisk (*) marks the three classifications operationalized in detail in Sections 4 through 6. Ten classifications are shown in eleven variants because the Altman et al. (2022) rule is reported in both calibrations. Family: ICR = Interest Coverage Ratio family; IRG = Interest Rate Gap family. Source: Refinitiv / LSEG; authors' calculations.

Three observations from the full set deserve emphasis. First, every one of the ten classifications outperforms the Nikkei 225 benchmark on both CAGR and Sharpe ratio. CAGR ranges from 5.1% (Storz et al.) to 10.8% (Banerjee and Hofmann 2018) against 3.8% for the benchmark; Sharpe ratios range from 0.33 to 0.54 against 0.28. The dispersion across classifications is narrow relative to the gap each opens against the benchmark. Second, the three classifications operationalized in the headline results were not selected as the highest-performing combinations. The best performer of the eleven variants is Banerjee and Hofmann (2018) (10.8% / 0.54), a non-headline row, and Nakamura (2023) (9.1% / 0.48), an IRG-family classification, exceeds two of the three headline classifications on CAGR. The three highlighted classifications were chosen to span the methodological range of the ICR family, not to maximize any single performance metric, and the strategy is therefore not the product of ex-post selection on the outcome variable. Third, the result is robust across the two principal signal families. The ICR family (classifications that read distress directly off the firm's

own interest-coverage ratio) delivers an average CAGR of 8.4% and an average Sharpe of 0.46 across its seven variants in Exhibit 7 (Altman is shown with both Z- and Z″-Score variants). The IRG family (classifications that infer distress from the gap between the firm's actual borrowing rate) delivers 7.1% and 0.41 across its four variants. The Matthews-correlation evidence in Exhibit 2 shows that the two families are essentially uncorrelated in the cross-section of firms they flag. The fact that two conceptually orthogonal classification traditions independently produce excess returns of the same order of magnitude is, in our view, the strongest single piece of evidence that the credit-mispricing premium is structural rather than an artifact of any one definition. Two further robustness exercises on the full set of ten classifications are reported in Appendix A. Widening the top-decile cut-off to the top 15 percent (Appendix A.2) leaves the qualitative conclusions unchanged, with average CAGR essentially flat (up by roughly 15 basis points) and six of the eleven variants posting a higher CAGR at the wider cut-off. The full Carhart four-factor decomposition for every classification (Appendix A.3) confirms positive annualized alpha throughout (range 0.2% to 5.2%), positive Value loadings for all eleven variants (range 0.058 to 0.408, statistically significant at the 1% level for 11 of them), and uniformly negative momentum loadings, mirroring the headline factor signature reported in Exhibit 6.

6.6 Is This More Than Investable Value? Placebo Portfolios and Spanning Tests

The Carhart value loadings of Section 6.3 invite an obvious objection: could the same returns be earned with a plain value sort? To answer it directly, we construct two placebo portfolios of identical construction (equally weighted top decile of the nominal index size, five-year formation averaging, annual rebalancing, net of 50 bp, financials excluded) sorted on a single valuation signal passed through the same fuzzy approach: low Tobin's Q, and high book-to-market. The placebo portfolios set a demanding benchmark: the Q sort earns a CAGR of 8.4% (Sharpe 0.46) and the book-to-market sort 5.0% (0.33), against 3.8% for the benchmark. A pure investable value sort in this universe therefore already earns a large share of what the zombie strategies earn. The economically interesting quantity is what remains after controlling for these investable value portfolios. Exhibit 8 reports spanning regressions of each strategy's daily excess returns on the placebo excess returns. The result is sharply two-sided. The IRG-family strategies and the most stringent distress screens are fully spanned: their alphas against both placebos jointly are statistically zero or negative. The ICR-family combinations that pair interest coverage with a market-value test retain positive alpha over both placebos jointly — 3.0% per annum for Banerjee and Hofmann (2018) and 2.3% for the 2020 variant (Newey–West t-statistics of 1.8 and 1.4), despite sharing 40–47% of their holdings with the pure-Q placebo. The incremental return of the credit screen is thus concentrated exactly where the survival evidence of Section 3.4 locates the mispricing: in classifications whose zombies show no elevated exit risk. Because ten classifications are examined in eleven classification variants, individual t-statistics overstate the strength of any single row, and three complementary statements discipline the inference. First, under a Bonferroni correction across all eleven variants, no individual spanning alpha remains significant at conventional levels, which is the appropriate reading of any single row in isolation. Second, the joint test of the two classifications singled out in advance in Section 2.5 — the coverage-plus-valuation rules of Banerjee and Hofmann, combined with equal weights — retains a spanning alpha of 2.7 percentage points per annum (t = 1.8).

Third, the same joint test applied to the full ICR family yields 1.3 points (t = 0.9), and to the IRG family +0.4 points (t = 0.3): the increment is a property of the coverage-plus-valuation configuration, not of the classification family at large. We accordingly describe the credit increment as economically meaningful but statistically moderate on its own; its cross-sectional signature, the concentration in classifications whose companies show no elevated exit risk, is the identifying evidence on which the mispricing interpretation rests. The fuzzy zombie strategy is best understood as a credit-quality screen that systematically upgrades an investable value portfolio; that the upgrade is delivered by the interest-coverage tradition and not by the interest-rate-gap tradition suggests that coverage-based classifications are preferable for investment purposes not only on data availability, as Section 2.5 argues, but on pricing content. The joint estimates are constructed transparently: each joint alpha is the arithmetic average of the corresponding Exhibit 8 estimates, and its standard error combines the Newey–West standard errors of Exhibit 8 with the correlation of the spanning-regression residuals (0.54 at the paper configuration; 0.42 to 0.57 across the alternative configurations of Appendix A).

Exhibit 8. Spanning Regressions on Investable Value Placebo Portfolios

Classification	CAGR	α vs Q	α vs B/M	α vs both	t (both)	R^2 (both)	Overlap w/ Q
Caballero et al. (2008)	5.6%	−0.97%	1.26%	−0.22%	−0.13	0.85	12%
Fukuda and Nakamura (2011)	5.8%	−1.36%	1.91%	−0.96%	−0.50	0.83	8%
Storz et al. (2017)	5.1%	−1.82%	1.17%	−1.34%	−0.62	0.80	22%
Adalet McGowan et al. (2017)	8.9%	1.40%	4.85% *	1.78%	0.98	0.85	14%
Banerjee and Hofmann (2018)	10.8%	2.75%	6.63% **	3.01% *	1.77	0.88	40%
Banerjee and Hofmann (2020)	10.2%	2.14%	6.21% **	2.31%	1.40	0.88	47%
Acharya et al. (2020)	7.9%	0.70%	3.87%	1.14%	0.57	0.82	6%
Favara et al. (2021)	7.2%	0.25%	3.41%	0.63%	0.30	0.79	6%
Altman et al. (2022) Z-Score	8.9%	1.74%	4.99% *	2.05%	1.14	0.83	13%
Altman et al. (2022) Z″-Score	7.5%	0.48%	3.68%	0.82%	0.44	0.81	13%
Nakamura (2023)	9.1%	1.44%	5.07% *	1.77%	0.86	0.84	23%

Note. Daily excess returns 2000–2025 regressed on placebo-portfolio excess returns; alphas annualized. Placebo portfolios: identical construction, sorted only on Tobin's Q (low) or book-to-market (high) through the same four-stage transformation. Overlap = average share of holdings in common with the Q placebo. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Newey–West standard errors use five lags; results are insensitive to lag choices between three and ten (t-statistics move by less than 0.04 across that range). Source: Refinitiv / LSEG; authors' calculations. Family-wise inference: Romano–Wolf stepdown-adjusted one-sided p-values across the eleven classifications are 0.22 (Banerjee and Hofmann 2018) and 0.35 (Banerjee and Hofmann 2020); see the closing paragraph of this section. Augmenting the spanning regressions with a daily size factor constructed inside the universe changes the alphas by less than 0.5 percentage points per annum, with size betas below 0.17 in absolute value, and size-matched placebo portfolios leave the ordering intact (Appendix A.8).

Two composition effects discipline the headline comparison. First, although the strategies hold no financial firms while the benchmark includes them, sector allocation explains essentially none of the gap: reconstructing the investable universe with and without financials over 2000–2025, the allocation effect of holding zero financials averages +0.04 percentage points per year on an equal-weighted basis, and the effect is negative in 11 of the 26 years. Second, an equally weighted portfolio of the full non-financial universe compounds at roughly 7.5% per annum over the sample, so part of the headline gap to the price-weighted Nikkei 225 reflects the equal weighting of index constituents rather than selection. The paper's central claim is accordingly a within-universe claim, and Exhibit 8 states it precisely: the top decile outperforms its own equally weighted opportunity set, and for the coverage-based headline rules a selection premium of 2.3 to 3.0 percentage points per annum survives value sorts of identical construction.

A final calculation disciplines Exhibit 8 as a whole. Because the exhibit reports eleven spanning alphas, its largest entry cannot be read as a stand-alone discovery (Harvey, Liu, and Zhu 2016). Applying the Romano and Wolf (2005) stepdown correction, based on a circular block bootstrap of the daily data with Newey–West studentization that preserves the substantial dependence across classifications (mean pairwise correlation of the centered statistics 0.42), raises the one-sided p-value of the strongest single alpha, Banerjee and Hofmann (2018), from 0.04 to 0.22; no classification clears conventional family-wise thresholds, and the adjusted values are insensitive to the block length. This quantifies a statement the paper has made throughout: the exhibit does not support selecting the best alpha after the fact. The inference the paper defends rests on the pre-specified test of Section 2.5, the interest-coverage-and-valuation subfamily as a single directional hypothesis (joint alpha 2.7 percentage points, t = 1.8, one-sided p of approximately 0.04), for which no selection across the exhibit occurs and no correction is due. The correction locates the evidential weight where the design placed it from the start.

7. Implications, Extensions, and Limitations

7.1 The Fuzzy Score as a Standalone Signal

The simplest implementation is the one tested in this paper: an equally weighted, long-only top-decile portfolio rebalanced annually on the Nikkei 225. The strategy is transparent, replicable, and operable on widely available data. It is also explicit about its risk exposure: a positive value loading and a negative momentum loading mean that the strategy will underperform in classic value-unfriendly regimes (sharp risk-off episodes, late-stage momentum rallies). An investor with a multi-year horizon and tolerance for tracking error against the benchmark in such regimes will find the historical reward-to-risk profile attractive. Appendix A.5 (Absolute Anchoring: The Depth of the Top Decile) adds an absolute yardstick to this profile, using thresholds the literature itself specifies — interest coverage of one, Tobin's Q of one. The top decile retains absolute depth in every regime: in 2020 through 2025, roughly a quarter of its names do not cover interest from operating earnings and some 95 percent trade below replacement value. The composition of that depth has shifted over the sample, from outright coverage distress in the early 2000s toward below-replacement valuations with moderately weak coverage against a much healthier universe; the premium of the recent decade is therefore earned on relative credit quality rather than on near-default distress.

7.2 Robustness on TOPIX and Under Capitalization Weighting

Two implementation variants complete the picture: the full point-in-time protocol re-run on the broader TOPIX universe, the panel on which Section 3 measures survival and zombie dynamics, and top-decile portfolios formed under capitalization weighting. The TOPIX cross-section is an order of magnitude wider than the Nikkei 225: the eligible formation universe (members with five consecutive score years) grows from roughly 650 names in 2000 to over 1,800 in 2025, so the top decile holds 64 to 185 names. Exhibit 9 (TOPIX Top-Decile Backtests: Excess Returns and Value-Placebo Spanning) reports the results for the pre-specified interest-coverage subfamily of Banerjee and Hofmann (2018, 2020). The absolute conclusion carries over with room to spare: the equal-

weighted subfamily average compounds at 9.5% per annum against 2.7% for the TOPIX price index, an annual excess of 7.5 percentage points (t = 2.5), and capitalization weighting delivers the same 9.5% (t = 2.8). The attribution, however, differs from the Nikkei 225 in an instructive way. Measured against an equally weighted portfolio of all TOPIX members, which compounds at 7.4% and embodies the pronounced small-firm effect of the Japanese cross-section, the equal-weighted excess narrows to 3.0 percentage points (t = 1.8). Against TOPIX-internal value placebos constructed under the Section 6.6 protocol, the subfamily alpha is −1.9 percentage points (t = −2.0) equal-weighted and 0.1 (t = 0.1) capitalization-weighted, with placebo R^2 of 0.95 and 0.87: on the broad universe, the zombie-decile premium is fully absorbed by generic value, whose own top deciles compound at 10.2% (Tobin's Q) and 11.1% (book-to-market). The same spanning test on the Nikkei 225 left a positive subfamily alpha of 2.7 percentage points (Exhibit 8, Spanning Regressions on Investable Value Placebo Portfolios). This contrast sharpens rather than weakens the credit-mispricing interpretation. Creditor forbearance runs through main-bank relationships that are strongest for the large, bank-connected firms of the blue-chip segment, and that is precisely where the premium survives value spanning; in the small-capitalization breadth of TOPIX, distress pricing collapses into the generic value effect. The TOPIX variant also matters beyond pricing: the survival evidence of Section 3 is measured on this panel, and it shows the classification rules behaving identically there. The Banerjee and Hofmann (2020) zombie cohort exits at the same rate as never-classified firms (exit ratio 1.00), so the mispricing mechanism is not an artifact of index membership.

Exhibit 9. TOPIX Top-Decile Backtests: Excess Returns and Value-Placebo Spanning

Portfolio	CAGR	Excess vs. index (pp)	t	Excess vs. universe (pp)	t	α vs. Q + B/M (pp)	t
Panel A. Equal-weighted							
Banerjee and Hofmann (2018)	9.3%	+7.3	2.48	+2.7	1.63	−2.2	−1.87
Banerjee and Hofmann (2020)	9.8%	+7.8	2.53	+3.2	1.79	−1.7	−1.67
Subfamily average	9.5%	+7.5	2.54	+3.0	1.80	−1.9	−1.98
Panel B. Capitalization-weighted							
Banerjee and Hofmann (2018)	8.2%	+6.7	2.59	+5.5	2.20	−0.8	−0.45
Banerjee and Hofmann (2020)	10.4%	+8.3	2.72	+7.2	2.51	+1.1	0.51
Subfamily average	9.5%	+7.6	2.80	+6.5	2.52	+0.1	0.08

Note. Top-decile portfolios on the TOPIX universe (index members ex financials with five consecutive score years), formed on the five-year moving average of the fuzzy score, rebalanced at each calendar year-end, delisted positions frozen at the last traded price through year-end, 50 bp transaction costs on turnover, formation years 2000–2025, daily constituent prices from LSEG/Refinitiv. Excess returns are annual arithmetic means over the TOPIX price index (CAGR 2.7%) and over an identically constructed portfolio of all TOPIX members under the same weighting scheme (CAGR 7.4% equal-weighted, 4.1% capitalization-weighted; the gap to the index reflects the small-firm effect and full-capitalization annual rebalancing). α vs. Q + B/M is the annualized intercept of a regression of daily returns on TOPIX-internal top-decile placebo portfolios formed on Tobin's Q and book-to-market alone, following the Section 6.6 protocol; those placebo deciles compound at 10.2% and 11.1% equal-weighted (9.0% and 9.4% capitalization-weighted). t-statistics are Newey–West. Subfamily = arithmetic average of the two Banerjee–Hofmann portfolios.

7.3 Limitations

Three limitations of the present study deserve explicit mention.

First, the headline investment universe is restricted to the Nikkei 225. For these large-capitalization names, capacity is not the binding concern: an equally weighted portfolio of 23 blue-chip stocks accommodates institutional mandates of substantial size. The capacity question arises in the TOPIX extension, where the broader universe adds small-capitalization names whose trading liquidity would constrain position sizes well before the signal loses resolution. A TOPIX implementation

therefore widens the opportunity set at the cost of a real capacity limit — the reverse of the usual trade-off. The pricing evidence of Section 7.2 adds a second qualification: what the added breadth delivers is largely generic value exposure rather than additional forbearance premium, so the capacity-constrained blue-chip segment is also where the distinctive part of the signal lives.

Second, the 50 bp transaction-cost assumption is generous in normal market conditions but may underestimate the round-trip cost of executing the strategy during episodes of severely impaired liquidity. The crisis-year drawdowns should be read with that caveat in mind.

Third, the Carhart factor model is a standard but blunt tool for attribution. A more granular factor decomposition (quality, profitability, low-volatility, or country-specific factors) could provide additional context. The credit-mispricing thesis would, however, predict the same qualitative signature: positive value, negative momentum, positive quality after recovery, with the alpha residual surviving these controls.

8. Conclusion

This paper has demonstrated, on a 26-year point-in-time backtest free of survivorship and look-ahead bias and net of a deliberately generous 50 bp transaction-cost charge, that the fundamentals from the academic zombie-firm literature can be repurposed as the basis of a systematic, transparent, fundamentally driven equity strategy on the Nikkei 225. The methodological centerpiece is the fuzzy score: rather than tagging each firm as either zombie or non-zombie, we convert the same accounting ratios that the literature uses for classification into a continuous score between zero and one. Every Nikkei 225 constituent receives a score in every year, the cross-section can be ranked end-to-end, and the top decile is always well-defined. The strategy that buys this top decile and rebalances annually delivers the headline results reported here. The economic intuition is credit mispricing. Investors apply excessive risk discounts to Nikkei 225 constituents that exhibit elevated zombie indicators (low interest coverage, low Tobin's Q, high leverage, weak sales growth, depressed Altman Z-Scores) even though the empirical recovery rate of these firms is materially higher than market pricing implies. The Kaplan-Meier survival evidence on the Tokyo Stock Exchange shows median zombie spells of one to two years across the classifications, annual exit intensities of at most 1.7%, and annual relapse probabilities of approximately 1% to 5.5%. The cumulative survival gap between zombies and non-zombies varies substantially across classifications: widest for the flow-based Storz et al. rule, intermediate for the Altman variants and Favara, and essentially absent for Banerjee and Hofmann (2020), whose cohort exits at the never-zombie rate — with the 2018 variant exiting at half that rate. The Cox proportional-hazards regressions show that higher capital intensity and stronger liquidity materially reduce the hazard of exit from the zombie pool — the classical collateral and forbearance channels. The combined message is unambiguous: recovery, not default, is the modal post-classification outcome, and the equity-market discount applied to these names is larger than realized credit losses justify. The resulting risk-premium gap is accessible through a long-only, equally weighted, top-decile portfolio rebalanced annually. Three findings deserve particular emphasis.

Across the three classifications operationalized in the headline results, the strategy delivers a CAGR of 7.2% to 10.2% against 3.8% for the Nikkei 225, an arithmetic mean return of 10.7% to 14.0% against 6.2%, and a Sharpe ratio of 0.41 to 0.52 against 0.28, all net of a 50 bp transaction-cost charge. The remaining seven classifications and the capitalization-weighted variants deliver qualitatively consistent results; all ten classifications outperform the Nikkei 225 benchmark (Section 6.5). On the broader TOPIX universe the top deciles beat the market index under both weighting schemes, though there the premium is largely absorbed by generic value; the part of the premium that survives value spanning is specific to the blue-chip segment (Section 7.2). The Nikkei 225 equal-weight implementation is reported in detail because it is the most directly investable specification and because equal weighting concentrates the score signal in the tails of the cross-section.

Carhart four-factor attribution confirms positive alpha (1.47% to 4.49% per annum), large and statistically significant value loadings (0.347 to 0.404), and negative momentum loadings: exactly the factor signature predicted by the credit-mispricing thesis. The excess return is anchored in the value effect and is not a disguised market or momentum bet. Spanning tests against investable value placebos discipline the interpretation: most of the return is shared with value, and the incremental 2.3–3.0 percentage points earned by the credit screen, economically meaningful but on its own statistically moderate ($t = 1.4$–1.8; jointly $t = 1.8$ for the pre-specified subfamily), sit precisely in the classifications whose zombies almost never exit, the sharpest version of the mispricing prediction, and one that generic value cannot generate. The TOPIX contrast completes the argument: on the broad universe generic value does span the zombie deciles, on the bank-connected blue-chip segment it does not.

The fuzzy score is investable across the cycle. In years with high binary zombie counts (post-1989 Japan, post-GFC Europe), the top decile concentrates in obviously distressed names. In years with low binary zombie counts (post-2013 Japan; current market conditions), the score still produces a complete cross-sectional ranking, and the top decile is still well-defined. The signal is a cardinal credit-quality ranking, not a binary count, and it survives the structural decline in zombification observed in many advanced economies in recent years — in Japan underpinned by the unwinding of the bank-centered crossholding system documented by Franks, Mayer, and Miyajima (2014). Beyond the specific strategy, the paper makes two broader contributions. First, it bridges a documented gap in the academic zombie-firm literature: that literature has been preoccupied with the count and the macro consequences of zombification and has paid limited attention to the cross-section of recoveries and defaults among classified firms. The credit-mispricing premium is the natural cross-sectional counterpart to those macro findings, and it is what an equity investor can capture. Second, the fuzzy scoring methodology is a general-purpose bridge from academic diagnosis to investment signal: any literature whose primary outputs are binary classifications of firms can be reread through the same four-step procedure to yield a continuous, rankable, replicable investment score.

Appendix A: Robustness Checks

This Appendix reports six robustness exercises on the fuzzy-score strategy presented in the main paper. Sections A.1 and A.2 vary the formation window and the portfolio cut-off; Section A.3 extends the Carhart decomposition to all ten classifications; Section A.4 reports the principal-component analyses summarized in Sections 3.2 and 4.5; Section A.5 reads the portfolios against absolute distress thresholds; Section A.6 varies the construction choices of the four-stage transformation one at a time; Section A.7 tabulates the survival, recovery, and Cox statistics behind Section 3; and Section A.8 controls the spanning tests of Exhibit 8 for within-universe size exposure. All implementations use the Nikkei 225, equal weighting, annual rebalancing, and are net of 50 basis points transaction costs unless otherwise stated.

A.1 Sensitivity to the Formation Period

The headline results use a five-year moving-average fuzzy score. Table A.1 reports CAGR (Panel A) and Sharpe ratios (Panel B) for formation windows of one through five years across all ten classifications. Two observations follow. First, every classification beats the Nikkei 225 benchmark at every formation window tested: the outperformance is not specific to the five-year specification. Second, the five-year baseline is never strictly the best-performing window (at printed precision it ties for best in two of the eleven variants) and lies at or below the median of the range for most classifications, a conservative choice. For the three classifications operationalized in detail (Banerjee and Hofmann 2020, Favara, Altman Z) the Sharpe ratio remains above the Nikkei 225 benchmark of 0.28 at every window.

Table A.1, Panel A. CAGR by Formation Period

Classification	5-yr MA (paper)	4-yr MA	3-yr MA	2-yr MA	1-yr MA
Nikkei 225 (Benchmark)	3.8%	3.8%	3.8%	3.8%	3.8%
Caballero et al. (2008)	5.6%	5.9%	6.5%	6.1%	5.0%
Fukuda and Nakamura (2011)	5.8%	6.4%	6.7%	7.7%	7.0%
Storz et al. (2017)	5.1%	6.0%	6.1%	8.0%	10.3%
Adalet McGowan et al. (2017)	8.9%	9.5%	9.2%	8.8%	8.5%
Banerjee and Hofmann (2018)	10.8%	9.7%	10.7%	10.8%	10.8%
Banerjee and Hofmann (2020)	10.2%	10.4%	11.6%	11.9%	10.1%
Acharya et al. (2020)	7.9%	7.7%	8.6%	7.3%	6.1%
Favara et al. (2021)	7.2%	7.2%	7.1%	6.7%	7.3%
Altman et al. (2022) Z-Score	8.9%	9.1%	8.6%	8.9%	8.9%
Altman et al. (2022) Z″-Score	7.5%	7.3%	7.6%	8.0%	7.3%
Nakamura (2023)	9.1%	9.1%	7.5%	7.2%	5.9%

Table A.1, Panel B. Sharpe Ratio by Formation Period

Classification	5-yr MA (paper)	4-yr MA	3-yr MA	2-yr MA	1-yr MA
Nikkei 225 (Benchmark)	0.28	0.28	0.28	0.28	0.28
Caballero et al. (2008)	0.36	0.37	0.39	0.38	0.33
Fukuda and Nakamura (2011)	0.36	0.38	0.39	0.43	0.41
Storz et al. (2017)	0.33	0.36	0.37	0.44	0.53
Adalet McGowan et al. (2017)	0.47	0.49	0.48	0.47	0.45
Banerjee and Hofmann (2018)	0.54	0.50	0.54	0.54	0.54
Banerjee and Hofmann (2020)	0.52	0.52	0.56	0.57	0.51
Acharya et al. (2020)	0.44	0.43	0.46	0.41	0.36
Favara et al. (2021)	0.41	0.41	0.41	0.39	0.41
Altman et al. (2022) Z-Score	0.49	0.50	0.48	0.49	0.48
Altman et al. (2022) Z″-Score	0.43	0.42	0.43	0.45	0.42
Nakamura (2023)	0.48	0.47	0.41	0.40	0.36

A.2 Sensitivity to the Portfolio Cut-Off

The headline portfolio is the top decile (10 percent) of the fuzzy score. Table A.2 widens the cut-off to the top 15 percent. The qualitative conclusions are unchanged: every row continues to beat the benchmark, average CAGR across the eleven variants is essentially flat (up by roughly 15 basis points), and six of the eleven variants post a higher CAGR at the wider cut-off. Widening the book dilutes the score tails only gradually.

Table A.2. Top Decile vs. Top 15 Percent

Classification	Top 10% (paper)	Top 15%
Nikkei 225 (Benchmark)	3.8%	3.8%
Caballero et al. (2008)	5.6%	6.2%
Fukuda and Nakamura (2011)	5.8%	6.4%
Storz et al. (2017)	5.1%	5.5%
Adalet McGowan et al. (2017)	8.9%	8.8%
Banerjee and Hofmann (2018)	10.8%	9.8%
Banerjee and Hofmann (2020)	10.2%	10.4%
Acharya et al. (2020)	7.9%	7.8%
Favara et al. (2021)	7.2%	8.6%
Altman et al. (2022) Z-Score	8.9%	8.6%
Altman et al. (2022) Z″-Score	7.5%	7.6%
Nakamura (2023)	9.1%	8.9%

A.3 Carhart Four-Factor Decomposition, All Ten Classifications

Table A.3 extends the Carhart four-factor regression of Exhibit 6 to the full set of ten classifications. The factor signature is uniform across the literature. Value loadings are positive for all eleven variants (range 0.058 to 0.408), statistically significant at the 1-percent level for 11 of them: the strategy is fundamentally a value strategy. Momentum loadings are uniformly negative (−0.151 to −0.073); the fuzzy ranking buys beaten-down stocks ahead of recovery. Annualized Alphas are positive throughout, ranging from 0.22% (Fukuda and Nakamura) to 5.15% (Banerjee and Hofmann 2018; the 2020 variant: 4.49%), economically meaningful but (under conservative Newey–West standard errors) short of statistical significance at conventional levels for all but one row. The Beta

clusters between 0.85 and 0.97 across the classifications and SMB loadings are positive for nine of the eleven variants (the Caballero and Storz rules are the exceptions), confirming that the result is a value tilt rather than a market-timing artifact.

Table A.3. Carhart Four-Factor Loadings, All Ten Classifications

Classification	Family	Alpha (ann.)	Beta	Size	Value	Momentum	R^2
Caballero et al. (2008)	IRG	3.41%	0.847	−0.158	0.058 ***	−0.120 ***	0.684
Fukuda and Nakamura (2011)	IRG	0.22%	0.910	0.198	0.408 ***	−0.149 ***	0.604
Storz et al. (2017)	ICR	2.65%	0.896	−0.069	0.086 ***	−0.079 ***	0.648
Adalet McGowan et al. (2017)	ICR	3.46%	0.942	0.209	0.380 ***	−0.151 ***	0.608
Banerjee and Hofmann (2018)	ICR	5.15% *	0.964	0.225	0.388 ***	−0.113 ***	0.604
Banerjee and Hofmann (2020)	ICR	4.49%	0.959	0.247	0.404 ***	−0.135 ***	0.593
Acharya et al. (2020)	IRG	2.16%	0.929	0.321	0.373 ***	−0.142 ***	0.589
Favara et al. (2021)	ICR	1.47%	0.895	0.289	0.391 ***	−0.110 **	0.567
Altman et al. (2022) Z-Score	ICR	3.50%	0.865	0.214	0.347 ***	−0.073 *	0.585
Altman et al. (2022) Z″-Score	ICR	2.12%	0.868	0.265	0.339 ***	−0.109 ***	0.570
Nakamura (2023)	IRG	3.64%	0.965	0.384	0.329 ***	−0.090 **	0.571

Note. Daily regressions 2000–2025; Newey–West standard errors; annualized alpha = daily alpha × 252. *** p < 0.01, ** p < 0.05, * p < 0.10. Source: Refinitiv / LSEG; authors' calculations.

A.4 Principal Components: Aggregation and Factor Structure

Sections 3.2 and 4.5 summarize two principal-component analyses; this section reports the details. Panel A of Table A.4 compares, for each classification, the fuzzy aggregation with its principal-component replacement: the share of variance explained by the first component of the classification's sub-scores, the Spearman correlation between the two formation rankings, and the differences in compound annual return and Sharpe ratio between the two portfolios, computed under identical conventions on matched samples. Panel B reports the loadings of the first two components of the eleven fuzzy scores themselves, estimated year by year on the complete-case panel.

Table A.4, Panel A. Fuzzy Aggregation versus First Principal Component, All Ten Classifications

Classification	PC1 var. share (%)	Rank corr.	ΔCAGR (pp)	ΔSharpe
Caballero et al. (2008)	100.0	1.00	+0.2	0.00
Fukuda and Nakamura (2011)	37.4	0.95	−0.4	−0.01
Storz et al. (2017)	49.2	0.82	+0.5	+0.02
Adalet McGowan et al. (2017)	54.8	0.78	+1.7	+0.04
Banerjee and Hofmann (2018)	50.4	0.92	+1.2	+0.05
Banerjee and Hofmann (2020)	74.6	0.99	−0.7	−0.02
Acharya et al. (2020)	54.6	0.88	−1.2	−0.03
Favara et al. (2021)	54.8	0.97	−1.1	−0.02
Altman et al. (2022) Z-Score	84.0	1.00	−0.8	−0.03
Altman et al. (2022) Z″-Score	79.8	1.00	+0.4	+0.02
Nakamura (2023)	37.1	0.30	−1.0	−0.08

Note. Loadings for Panel A are estimated once per formation year on the pooled panel of the five formation score-years, so the procedure remains strictly point-in-time. Rank corr. = Spearman correlation between the fuzzy and the principal-component formation rankings, averaged over portfolio years. ΔCAGR and ΔSharpe = principal-component portfolio minus fuzzy portfolio on matched samples (Storz et al. from 2005, where complete moving-average inputs first become available). Nakamura's fuzzy-OR aggregation is intrinsically nonlinear, and its low rank correlation reflects the absence of a linear counterpart rather than estimation noise. Source: Refinitiv / LSEG; authors' calculations.

Table A.4, Panel B. Factor Structure of the Eleven Fuzzy Scores

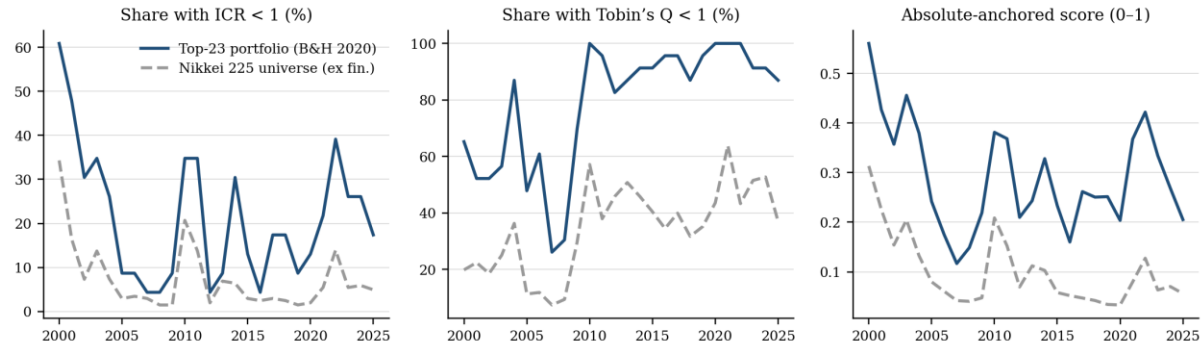
Classification	Family	PC1 loading	PC2 loading
Caballero et al. (2008)	IRG	0.10	+0.60
Fukuda and Nakamura (2011)	IRG	0.33	+0.31
Storz et al. (2017)	ICR	0.28	−0.06
Adalet McGowan et al. (2017)	ICR	0.27	−0.25
Banerjee and Hofmann (2018)	ICR	0.31	−0.21
Banerjee and Hofmann (2020)	ICR	0.34	−0.13
Acharya et al. (2020)	ICR/IRG	0.35	+0.23
Favara et al. (2021)	ICR	0.33	−0.05
Altman et al. (2022) Z-Score	ICR	0.37	−0.15
Altman et al. (2022) Z″-Score	ICR	0.35	−0.14
Nakamura (2023)	IRG	0.13	+0.55

Note. Cross-sectional principal components of the eleven fuzzy scores on the complete-case Nikkei 225 panel, estimated year by year, 2000–2024; loadings averaged over years; the second component is oriented so that the interest-rate-gap block loads positively. Average share of variance: PC1 55.0%, PC2 19.2%, PC3 8.7%; average adjacent eigenvalue ratios 2.9, 2.3, and 1.6 (Ahn and Horenstein 2013). Source: Refinitiv / LSEG; authors' calculations.

A.5 Absolute Anchoring: The Depth of the Top Decile

The four-stage transformation is deliberately relative: every input is standardized within the year, so a score locates a firm against its contemporaries and is silent on how distressed the firm — or the market — is in absolute terms. As a complement, this section anchors the key inputs at the absolute thresholds the literature itself specifies: interest coverage of one (interest no longer covered by operating earnings), an interest-rate gap of zero (borrowing below the safe reference rate), Tobin's Q of one (market value below asset replacement), and Altman's distress boundary of 1.81. No parameter is estimated. Figure A.1 plots, for each portfolio year, the share of firms below the coverage and valuation anchors in the investable universe and in the Banerjee and Hofmann (2020) portfolio, together with an absolute-anchored composite score; Table A.5 reports sub-period averages together with medians of the underlying ratios, taken from the last pre-formation fiscal year.

Figure A.1. Absolute Distress Depth of the Investable Universe and the Top-Decile Portfolio, 2000–2025



Note. Solid line: Banerjee and Hofmann (2020) portfolio; dashed line: Nikkei 225 investable universe excluding financials. Anchors: ICR = 1; Tobin's Q = 1. The absolute-anchored score is the geometric mean of fixed logistic maps centered at the two anchors. Source: Refinitiv / LSEG; authors' calculations.

Table A.5. Absolute Depth by Sub-Period: Threshold Shares and Median Ratios

Period	Group	% ICR<1	% Q<1	% Z<1.81	Abs. score	Med. ICR	Med. Q	Med. leverage
2000–04	Portfolio	40.0	62.6	67.0	0.44	1.4	0.97	0.54
	Universe	15.8	24.4	48.4	0.21	4.3	1.09	0.43
2005–09	Portfolio	7.0	47.0	80.9	0.18	4.2	1.03	0.52
	Universe	2.5	13.9	33.7	0.05	14.0	1.20	0.33
2010–14	Portfolio	22.6	91.3	86.1	0.31	3.2	0.91	0.46
	Universe	10.0	47.5	44.9	0.13	9.0	1.01	0.33
2015–19	Portfolio	12.2	93.0	80.9	0.23	4.3	0.88	0.42
	Universe	2.5	36.3	34.6	0.05	21.5	1.08	0.23
2020–25	Portfolio	23.9	94.9	83.3	0.30	3.8	0.85	0.39
	Universe	6.3	48.6	33.5	0.07	23.2	1.01	0.23

Note. Sub-period averages of annual values; Portfolio = Banerjee and Hofmann (2020) selection at formation; Universe = Nikkei 225 constituents excluding financials; variables from the last pre-formation fiscal year. The share of firm-years with a negative interest-rate gap is between zero and one percent in the universe and zero in the portfolios in every sub-period and is omitted from the table: near-zero policy rates compress the rate dispersion that the criterion requires, so the interest-rate-gap rules operate in this universe entirely through their relative reading. Source: Refinitiv / LSEG; authors' calculations.

Three facts stand out. First, the de-zombification of corporate Japan is fully visible in the absolute series: the universe median coverage ratio rises from 4.3 to 23, median leverage falls from 0.43 to 0.23, and the universe share below the coverage anchor drops from 16 to 6 percent. Part of the coverage improvement is mechanical — near-zero policy rates shrink interest expense — but the parallel deleveraging indicates that most of it is real. Second, the portfolio retains absolute depth in every regime, holding a persistent multiple of the universe's share of firms with uncovered interest. Third, the composition of that depth shifts over the sample, from outright coverage distress in the early years toward below-replacement valuations with moderately weak coverage; Section 7.1 draws the implication for how the premium of the recent decade should be read.

A.6 Construction Choices and Input Noise

Section A.4 replaces the aggregation stages of the four-stage transformation wholesale; this section varies its construction choices one at a time, holding everything else fixed: the winsorization bounds (0.5/99.5, 2.5/97.5, and no winsorization, against the baseline 1/99), the signed-log rule (applied always and never, against the |skewness| > 0.5 trigger), the steepness of the logistic map (one half and twice the baseline slope), and the aggregation operator (arithmetic mean and the minimum — the hard fuzzy-AND, the most literal translation of the binary definitions' requirement that every condition bind simultaneously — against the geometric mean). Table A.6 reports, for the three headline classifications, the change in compound annual return against the identically constructed baseline, the Spearman correlation of the formation ranking with the baseline ranking, and the average overlap of the selected portfolio; Sharpe ratios move by at most 0.05 across all variants.

Table A.6. Construction Choices, One at a Time

Variant	ΔCAGR	ρ	Ovl. %	ΔCAGR	ρ	Ovl. %	ΔCAGR	ρ	Ovl. %
Baseline (paper)	0.0	1.00	100	0.0	1.00	100	0.0	1.00	100
Winsorization 0.5/99.5	−0.2	1.00	98	−0.2	1.00	97	−0.3	1.00	100
Winsorization 2.5/97.5	+0.2	1.00	97	+0.2	1.00	97	+0.3	1.00	99
No winsorization	+0.0	1.00	96	+0.1	1.00	96	−0.5	1.00	99
Signed-log always	0.0	1.00	100	+0.2	1.00	97	0.0	1.00	100
Signed-log never	−2.6	0.87	58	+0.1	0.98	86	−1.1	0.98	88
Logistic slope × ½	−0.4	1.00	94	+0.1	1.00	93	−0.9	1.00	98
Logistic slope × 2	−0.3	1.00	93	+0.7	1.00	89	+0.0	1.00	96
Arithmetic mean	−0.4	1.00	96	−1.0	0.99	91	−0.3	1.00	98
Minimum (fuzzy-AND)	−0.7	0.97	78	+1.1	0.95	76	+0.4	0.99	81

Note. Column blocks: Banerjee and Hofmann (2020), Favara et al. (2021), Altman et al. (2022) Z-Score. ΔCAGR in percentage points against the identically constructed baseline; ρ = Spearman correlation of the formation ranking with the baseline ranking, averaged over portfolio years; Ovl. = average share of holdings in common with the baseline portfolio. Every variant portfolio beats the Nikkei 225 benchmark. Source: Refinitiv / LSEG; authors' calculations.

Two results carry the table. First, the winsorization bounds, the steepness of the logistic map, and the choice between geometric and arithmetic averaging matter very little: rank correlations of 0.99 and above, portfolio overlaps of 89 to 100 percent, and changes in compound return below one percentage point. The hard fuzzy-AND builds the most different portfolios of the grid (76 to 81 percent overlap) yet performs as well or better, so the economic conclusion does not depend on the softness of the geometric mean. Second, the one component that materially matters is the signed-log transformation: removing it entirely cuts the Banerjee and Hofmann (2020) overlap to 58 percent, because the heavily skewed raw coverage ratio then dominates the cross-sectional standardization and crowds out the valuation signal — while applying the transformation always is indistinguishable from the baseline, since the skewness trigger fires almost everywhere. The transformation itself is load-bearing; its threshold is a matter of convention. Even the no-transformation variant beats the benchmark by more than five percentage points per annum.

A second exercise changes the inputs rather than the construction: multiplicative Gaussian noise of two and five percent on the raw fundamental inputs, one hundred bootstrap draws per setting. At two percent — a realistic magnitude for restatements and vendor differences — the selected portfolios overlap the noise-free baseline at 96.5 to 99.1 percent, i.e., fewer than one exchanged name per year; at five percent the overlap remains between 92.2 and 97.8 percent, at most two names. Across twenty full backtest draws at two percent noise, the Banerjee and Hofmann (2020) compound annual return moves with a standard deviation of 0.3 percentage points around its baseline. The selection does not hinge on borderline names.

A.7 Survival and Recovery by Classification

Sections 3.3 through 3.5 summarize the dynamics of zombie status on the TOPIX panel; this section tabulates the underlying statistics. Panel A reports, for each classification, the population of the zombie pool, spell durations, annual exit, recovery, and relapse intensities, and the Kaplan–Meier survival gaps between the zombie and never-zombie cohorts at the five- and ten-year horizons. Panel B reports the pooled Cox proportional-hazards estimates for the six classifications with at least fifty observed exits. Exit is delisting of any cause, dated from the vendor's identifier convention with rename correction; cohorts are formed at first classification, and censoring is at the end of the price panel (May 2026).

Table A.7, Panel A. Zombie Dynamics, Exit Intensities, and Survival Gaps

Classification	Zombie firm-years	Cohort firms	Median spell (yrs)	Exit % p.a.	Exit ratio	Recovery % p.a.	Relapse % p.a.	KM gap 5y / 10y (pp)
Caballero et al. (2008)	320	158	1	0.0	0.00	55.1	1.1	0.5 / −0.7
Fukuda and Nakamura (2011)	5	—	1	0.0	0.00	—	1.7	—
Storz et al. (2017)	289	201	1	1.7	2.91	72.3	1.6	6.9 / 10.5
Adalet McGowan et al. (2017)	1,357	493	2	0.7	1.24	39.0	1.8	2.2 / 1.8
Banerjee and Hofmann (2018)	634	256	1	0.3	0.52	43.6	1.3	2.3 / 2.6
Banerjee and Hofmann (2020)	1,166	499	1	0.6	1.00	53.6	2.6	3.1 / 1.7
Acharya et al. (2020)	38	—	1	0.0	0.00	56.8	0.8	—
Favara et al. (2021)	1,233	547	1	1.1	1.79	57.6	3.1	3.0 / 2.4
Altman et al. (2022) Z-Score	1,211	431	2	1.2	1.97	36.4	1.2	3.8 / 4.2
Altman et al. (2022) Z″-Score	1,623	502	2	0.9	1.57	32.8	1.6	2.9 / 3.5
Nakamura (2023)	2,786	1,231	1	1.0	1.67	73.0	5.5	2.7 / 2.8

Note. TOPIX panel excluding financials, classification flags FY1995–FY2023, exits through May 2026. Zombie firm-years = classified firm-years; cohort firms = firms from first classification (Kaplan–Meier cohort); the never-zombie exit intensity is approximately 0.6% per annum throughout, and the exit ratio divides the zombie by the never-zombie intensity. Recovery = annual probability of leaving the zombie state; relapse = annual probability that a previously classified, currently non-classified firm re-enters the pool. KM gap = never-zombie minus zombie cumulative survival at five and ten years. The Fukuda–Nakamura and Acharya pools are too thin for cohort analysis (5 and 38 firm-years): near-zero policy rates compress the rate dispersion their interest-rate-gap criterion requires. Exit is delisting of any cause and includes takeovers. Source: Refinitiv / LSEG; authors' calculations.

Table A.7, Panel B. Pooled Cox Proportional Hazards in the Zombie Cohort

Classification	Exits	Asset turnover	Current ratio	Capital intensity	Leverage	Market share
Adalet McGowan et al. (2017)	64	1.15	0.82	0.86	0.92	0.95
Banerjee and Hofmann (2020)	71	0.98	1.24	0.87	1.09	1.00
Favara et al. (2021)	66	0.95	0.74	0.79	0.86	1.06
Altman et al. (2022) Z-Score	67	0.84	0.70	0.73	0.94	0.97
Altman et al. (2022) Z″-Score	82	0.85	0.69	0.72	0.91	1.07
Nakamura (2023)	193	0.98	0.83	0.84	0.99	0.86

Note. Hazard ratios per one within-year standard deviation of the covariate; covariates winsorized at the 1st and 99th percentiles, signed-log transformed where skewed, z-standardized within year; one pooled model per classification in the zombie cohort, restricted to classifications with at least fifty observed exits. A hazard ratio below one is protective (lower exit hazard). Source: Refinitiv / LSEG; authors' calculations.

A.8 Size Control in the Spanning Tests

A natural question after the Carhart attribution of Exhibit 6 is whether part of the spanning residual of Exhibit 8 reflects within-universe size exposure rather than the credit screen: the headline strategies carry positive size loadings, and the value placebos control cheapness but not size separately. Two size controls answer it. First, a daily size factor is constructed inside the universe — at each formation date, constituents are split at the median market capitalization, and the factor is the return difference between the equally weighted small and large halves, annually rebalanced under the delisting convention of Section 5.2 — and added to the spanning regressions. Second, the value placebos are re-selected within size halves so that each year's placebo matches the size composition of the Banerjee and Hofmann (2020) selection. Table A.8 reports both controls; estimates are replication-pipeline values (the convention of the Romano–Wolf adjustment in Section 6.6), so the baseline column reproduces the replication package rather than the production values of Exhibit 8, within the disclosed tolerances. The answer is unambiguous: the size betas of the spanning residuals are below 0.17 in absolute value for all eleven variants, adding the size factor changes no alpha by more than 0.5 percentage points per annum, and size-matched placebos leave the ordering intact. The increment measured in Exhibit 8 is not a size effect.

Table A.8. Size-Controlled Spanning: SMB-Augmented Regressions and Size-Matched Placebos

Classification	α vs Q+B/M	t	α +SMB	t	β SMB	α size-matched	t
Caballero et al. (2008)	−0.74	−0.45	−0.32	−0.19	−0.156	−0.81	−0.48
Fukuda and Nakamura (2011)	−0.12	−0.07	−0.18	−0.10	0.021	−0.51	−0.28
Storz et al. (2017)	1.11	0.57	0.92	0.47	0.167	0.22	0.11
Adalet McGowan et al. (2017)	−0.28	−0.17	−0.43	−0.25	0.053	−0.67	−0.39
Banerjee and Hofmann (2018)	−0.27	−0.19	−0.59	−0.41	0.117	−0.60	−0.42
Banerjee and Hofmann (2020)	2.10	1.65	2.00	1.57	0.034	1.66	1.28
Acharya et al. (2020)	−0.60	−0.33	−0.86	−0.48	0.094	−0.99	−0.54
Favara et al. (2021)	−0.17	−0.10	−0.55	−0.31	0.137	−0.56	−0.31
Altman et al. (2022) Z-Score	0.98	0.57	1.18	0.68	−0.071	0.57	0.32
Altman et al. (2022) Z″-Score	−1.27	−0.71	−1.42	−0.79	0.054	−1.58	−0.86
Nakamura (2023)	−0.17	−0.10	0.24	0.14	−0.151	−0.26	−0.15

Note. Replication-pipeline estimates on the 30 June 2026 pull (provenance as in the Romano–Wolf adjustment of Section 6.6; baseline alphas therefore differ from the production values of Exhibit 8 within the disclosed tolerances — see the replication package, TOLERANCES.md). α = annualized intercept (percentage points per annum) of a regression of daily strategy returns on both placebo returns, without and with the internal size factor, and on the size-matched placebos; Newey–West standard errors with five lags. β SMB = loading on the internal size factor. Source: Refinitiv / LSEG; authors' calculations.

References

Acharya, V. V., M. Crosignani, T. Eisert, and C. Eufinger. 2020. “Zombie Credit and (Dis-)Inflation: Evidence from Europe.” SSRN Electronic Journal.

Acharya, V. V., T. Eisert, C. Eufinger, and C. Hirsch. 2019. “Whatever It Takes: The Real Effects of Unconventional Monetary Policy.” Review of Financial Studies 32 (9): 3366–3411.

Adalet McGowan, M., D. Andrews, and V. Millot. 2017. “The Walking Dead?: Zombie Firms and Productivity Performance in OECD Countries.” OECD Economics Department Working Papers 1372.

Ahearne, A. G., and N. Shinada. 2005. “Zombie Firms and Economic Stagnation in Japan.” International Economics and Economic Policy 2 (4): 363-381.

Ahn, S. C., and A. R. Horenstein. 2013. “Eigenvalue Ratio Test for the Number of Factors.” Econometrica 81 (3): 1203–1227.

Altman, E. I. 1968. “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy.” Journal of Finance 23 (4): 589-609.

Altman, E. I., R. Dai, and W. Wang. 2022. “Global Zombies.” SSRN Electronic Journal.

Banerjee, R., and B. Hofmann. 2018. “The Rise of Zombie Firms: Causes and Consequences.” BIS Quarterly Review, September.

Banerjee, R., and B. Hofmann. 2020. “Corporate Zombies: Anatomy and Life Cycle.” SSRN Electronic Journal.

Brown, S. J., W. N. Goetzmann, and S. A. Ross. 1995. “Survival.” Journal of Finance 50 (3): 853–873.

Caballero, R. J., T. Hoshi, and A. K. Kashyap. 2008. “Zombie Lending and Depressed Restructuring in Japan.” American Economic Review 98 (5): 1943-1977.

Carhart, M. M. 1997. “On Persistence in Mutual Fund Performance.” Journal of Finance 52 (1): 57-82.

Connor, G., and R. A. Korajczyk. 1986. “Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis.” Journal of Financial Economics 15 (3): 373–394.

Cox, D. R. 1972. “Regression Models and Life-Tables.” Journal of the Royal Statistical Society, Series B 34 (2): 187–220.

Elinger, S. B. 2025. “Zombification and Investment Opportunities in the Japanese Stock Market.” Master’s Thesis, University of St. Gallen.

Fama, E. F., and K. R. French. 1993. “Common Risk Factors in the Returns on Stocks and Bonds.” Journal of Financial Economics 33 (1): 3-56.

Favara, G., C. Minoiu, and A. Perez-Orive. 2021. “U.S. Zombie Firms: How Many and How Consequential?” FEDS Notes.

Franks, J. R., C. Mayer, and H. Miyajima. 2014. “The Ownership of Japanese Corporations in the 20th Century.” Review of Financial Studies 27 (9): 2580–2625.

French, K. R. n.d. “Data Library.” Tuck School of Business at Dartmouth. Accessed June 2026.

Fukuda, S., and J. Nakamura. 2011. “Why Did ‘Zombie’ Firms Recover in Japan?” World Economy 34 (7): 1124-1137.

Goetzmann, W. N., and P. Jorion. 1999. “Re-Emerging Markets.” Journal of Financial and Quantitative Analysis 34 (1): 1–32.

Graham, B., and D. L. Dodd. 1934. Security Analysis. New York: McGraw-Hill.

Harvey, C. R., Y. Liu, and H. Zhu. 2016. “… and the Cross-Section of Expected Returns.” Review of Financial Studies 29 (1): 5–68.

Hoshi, T. 2006. “Economics of the Living Dead.” Japanese Economic Review 57 (1): 30-49.

Kaplan, E. L., and P. Meier. 1958. “Nonparametric Estimation from Incomplete Observations.” Journal of the American Statistical Association 53 (282): 457–481.

Lettau, M., and M. Pelger. 2020. “Factors That Fit the Time Series and Cross-Section of Stock Returns.” Review of Financial Studies 33 (5): 2274–2325.

Matthews, B. W. 1975. “Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme.” Biochimica et Biophysica Acta — Protein Structure 405 (2): 442–451.

Mingarelli, L., B. Ravanetti, T. Shakir, and J. Wendelborn. 2022. “Dawn of the (Half) Dead: The Twisted World of Zombie Identification.” ECB Working Paper Series 2743, European Central Bank.

Nakamura, J. 2023. “A 50-Year History of ‘Zombie Firms’ in Japan.” Japan and the World Economy 66: 101188.

Newey, W. K., and K. D. West. 1987. “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” Econometrica 55 (3): 703-708.

Peek, J., and E. S. Rosengren. 2005. “Unnatural Selection: Perverse Incentives and the Misallocation of Credit in Japan.” American Economic Review 95 (4): 1144-1166.

Romano, J. P., and M. Wolf. 2005. “Stepwise Multiple Testing as Formalized Data Snooping.” Econometrica 73 (4): 1237–1282.

Schumpeter, J. A. 1942. Capitalism, Socialism and Democracy. New York: Harper & Brothers.

Storz, M., M. Koetter, R. Setzer, and A. Westphal. 2017. “Do We Want These Two to Tango?: On Zombie Firms and Stressed Banks in Europe.” ECB Working Paper Series 2104.

Tobin, J. 1969. “A General Equilibrium Approach to Monetary Theory.” Journal of Money, Credit and Banking 1 (1): 15-29.

The authors used AI-assisted tools (Claude, Anthropic) for drafting, editing, and data synthesis. All analytical content, interpretations, and conclusions are the sole responsibility of the authors.