The Agony and the Ecstasy

Constructing a "Crash-Filtered" Equity Index using Machine Learning

Tristan Leiter

Vienna University of Business and Economics

January 28, 2026

The "Agony and Ecstasy" of Indexing

- **The Passive Investing Dilemma:** Indices capture the "Ecstasy" of extreme winners but systematically force investors to hold the "Agony" of imploding stocks.
  - Cembalest (2014) finds that equity indices are driven by a small tail of winners, while approx. 40% of constituents suffer "catastrophic declines" ($>$ 70% drawdown) without recovery.
- Tewari et al. (2024) formalize this as a **Catastrophic Stock Implosion (CSI)** .
  - It is a distinct event: a severe price downturn followed by a "zombie" period of prolonged stagnation.
- Passive investing captures the winners, but fails to filter the "Agony" until it is too late. Standard metrics fail to distinguish between **recoverable volatility** and **terminal implosion** .

Research Question

Main Research Question:

To what extent does a 'Crash-Filtered' equity index, constructed via probabilistic implosion modeling, generate superior risk-adjusted returns compared to the market benchmark and traditional minimum-volatility strategies?

Subquestions:

- Does the integration of latent features derived from Denoising Autoencoders significantly improve the predictive performance (Average Precision) of Ensemble models compared to those trained solely on raw financial data?
- Do Ensemble methods exhibit a superior ability to distinguish between recoverable volatility ('Ecstasy') and terminal implosion ('Agony') compared to indiscriminate volatility-based exclusion strategies?
- Does shortening the prediction horizon from 12 months to 6 months significantly enhance the index's ability to react to distress signals, or does it introduce excessive turnover without performance gains?

Limits of Traditional Models

- **The "Quality Trap":** Perceived safety signals can be misleading. Penman and Reggiani (2018) suggest that low B/P ratios often reflect uncertainty rather than value, while profitability measures lose predictive power over long horizons .

- **The False Positive Dilemma:** Traditional bankruptcy models (e.g., Altman (1968)) and risk scaling strategies, like Minimum-Volatility (MinVol.), fail to distinguish between 'good' volatility (growth) and 'bad' volatility (implosion), systematically excluding winners.

- **Contribution: A "Crash-Filtered" Index**
  Moving from "volatility forecasting" to "probabilistic implosion modeling":
  - This thesis proposes a Machine-Learning model to predict the "Agony" stocks whilst preserving the "Ecstasy" candidates.
  - The model predictions will be used to create a crash-filtered index. Its performance will be compared to the market benchmark and traditional MinVol. strategies.

Methodology I: Dependent Variable

Goal: Set up the dependent variable Catastrophic Stock Implosion (CSI)
Following Tewari et al. (2024), a stock is classified as a CSI ($y = 1$) if it satisfies:

- **Initial Crash (C):** $> 80\%$ drawdown from trailing peak (beginning of the "zombie" period).
- **Non-Recovery:** A maximum cumulative return of -20% in the "zombie" period.
- **Zombie Period:** Duration of 1.5 years.

Tewari et al. (2024) employ a yearly prediction horizon ($h = 12$ months), which this thesis follows:

$$y_{i,t} = \begin{cases} 1 & \text{if stock } i \text{ triggers } C = 80\% \text{ within } [t, t+h] \text{ and zombie criteria met,} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Methodology II: Modelling

- Using an **Autoencoder** to compress the noisy financial data into latent "distress structures" that linear PCA misses.
- **Supervised Learning:** Training ensemble-methods (Random Forest, XGBoost, CatBoost, LightGBM) on both raw data and latent features to predict the probability of CSI.

Cross-Validation optimization will be conducted via **Average Precision (AP)**. Since the dataset is imbalanced, AP provides for a more robust signal for hyperparameter tuning without committing to a specific decision threshold.

The models will be evaluated based on **Recall at fixed FPR**.
A constraint-based metric aligns more closely with the practical "Risk-Budget" in Portfolio Management. The objective is to maximize the number of Agony stocks identified (Recall) whilst capping the exclusion of Ecstasy stocks.

Methodology III: Index Construction

To ensure a proper **Out-of-Sample test**, the dataset will be split into three parts:

- Training-Set for Cross-Validation.
- Test-set for Model-selection
- Out-of-Sample set for backtesting.

The best-performing model will within the test-set be used to predict a CSI for the consecutive year. The "Crash-Filtered" Index will be rebalanced anually at the end of each calendar year. The probability threshold, θ, is selected based on the desired FPR rate (for example 3% or 5%).

Index Construction

The "Crash-Filtered" Index systematically excludes constituents where the predicted probability $\hat{p}_{CSI} > \theta$. Unlike standard classifiers that use a default $\theta = 0.5$, this threshold is **dynamically calibrated** to satisfy a specific risk constraint (e.g., $FPR \leq 5\%$), ensuring the exclusion rate aligns with the investor's "budget" for opportunity cost.

CRSP

The Universe (CRSP)

- **Scope:** US Common Equities (NYSE, AMEX, NASDAQ).
- **Timeline:** 1998 - 2024.
- **Size:** $3,255$ unique firms, 46,294 firm-year observations.
- **Constraint:** Minimum listing lifetime of 5 years to ensure sufficient learning history (13 years on average per firm).
- **CSI-Events:** $2,507$ CSIs in Total.

Category	Cohort Distribution	
	Imploded Firms	Never Imploded
High Growth ($> 10\%$)	12.5%	42.6%
Moderate Growth ($5 - 10\%$)	16.4%	24.2%
Low Growth ($2 - 5\%$)	10.1%	7.7%
Stagnation ($-2 - 2\%$)	0.0%	0.1%
Value Destruction (Intermitent Recovery)	48.5%	16.8%
Value Destruction (No Recovery)	1.4%	1.1%
Unclassified	11.1%	7.6%

Note: Categories are measured by the average geometric return.

Compustat

Features

- **Accounting:** Balance-Sheet and Income-Statement variables.
- **Macro:** Variables with possible interactions with accounting variables (interest rate, others). Obtained from the FRED database.
- **Other:** Specifically targeting "Zombie" precursors:
  - *Employees* (Number of employees).
  - *Rental expenses*

Compustat

Category	Variable Code	Description
Balance Sheet: Assets	`at` / `act`	Total Assets / Total Current Assets
	`che` / `ivst`	Cash & Short-Term Inv. / Short-Term Investments
	`rect` / `invt`	Receivables (Channel stuffing risk) / Inventories
	`wcap`	Working Capital (Liquidity buffer)
	`ppent` / `intan`	Net PP&E / Intangibles (Soft assets)
	`gdwl`	Goodwill (Impairment risk)
	`txdba`	Deferred Tax Asset (Long Term) - *Proxy for NOLs*
Balance Sheet: Liab/Eq	`lt` / `lct`	Total Liabilities / Total Current Liabilities
	`dltt` / `dlc`	Long-Term Debt / Debt in Current Liabilities
	`dd1`	Long-Term Debt Due in 1 Year (*Refinancing wall*)
	`ap` / `txp`	Accounts Payable / Income Taxes Payable
	`txditc`	Deferred Taxes & Inv. Tax Credit (Non-current)
	`seq` / `re`	Stockholders' Equity / Retained Earnings (*Accum. Deficit*)
	`pstk` / `mib`	Preferred Stock / Noncontrolling Interest
	`tstk`	Treasury Stock (Contra-equity)

Expected Contribution

- **Replication:** Confirming the methodology and results of the working paper by Tewari et al. (2024) with the CRSP and Compustat Data.
- **Extension:** Extending the results of Tewari et al. (2024) by including autoencoders for extraction of the informational content in the features and by using the model-predictions for index-constructions.
- **Challenging traditional risk-scaling approaches:** Showing that ML-applications are better suited to distinguish between Ecstasy and Agony stocks compared to classical risk-scaling methods, like Low-Volatility, Low-Beta or Altman's Z-score. Autoencoders and ensemble-methods are well suited for capturing the complex interactions between accounting and macro variables.

Bibliography

Edward I. Altman. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4):589–609, 1968.

Michael Cembalest. The agony & the ecstasy: The risks and rewards of a concentrated stock position. Eye on the market special edition, J.P. Morgan Asset Management, 2014. September 2, 2014.

Stephen Penman and Francesco Reggiani. Fundamentals of value versus growth investing and an explanation for the value trap. *Financial Analysts Journal*, 74 (4):103–119, 2018. doi: 10.2469/faj.v74.n4.6.

Zaki Tewari, Michal Galas, and Philip Treleaven. Predicting catastrophic stock implosions: A machine learning approach. Technical report, University College London, 2024. Working Paper.