Vienna University of Economics and Business

Master Thesis Research Proposal

The Agony and the Ecstasy: Constructing a "Crash-Filtered" Equity Index using Machine Learning

Author:
Tristan Leiter

Submission Date:
January 21, 2026

Supervisor:
Univ.Prof. Dr. Kurt Hornik
Department of Finance, Accounting and Statistics

Chapter 1

Problem Description

1.1 Introduction

Research by J.P. Morgan Asset Management highlights an empirical phenomenon referred to as "The Agony and the Ecstasy" (Cembalest, 2014, 2024) revealing that equity indices, like the Russell 3000, are overwhelmingly influenced by extreme stock performances. While a small percentage of winners contribute the vast majority of excess returns, approximately 40% of all constituents suffer a "catastrophic decline," which is defined as a drawdown of 70% or more from their peak without a subsequent recovery.

Recent literature formalizes this phenomenon as a **"Catastrophic Stock Implosion"** (CSI) (Tewari et al., 2024). Unlike standard volatility, an implosion represents a distinct market event characterized by a severe price downturn followed by prolonged stagnation and minimal probability of recovery. This presents a critical challenge for index construction: passive investing captures the "Ecstasy" of winners but systematically forces investors to hold the "Agony" of these imploding assets (at least until removal from the constituent index). Crucially, as explored by Cembalest (2014, 2024), these declines are not limited to speculative "junk" companies; a significant portion of losers display "healthy" fundamentals prior to collapse, suggesting traditional metrics fail to capture the dynamics preceding permanent capital decline.

1.2 Status Quo

1. **The Quality Trap:** Recent literature has explored indicators which are perceived to be synonymous with healthy stock fundamentals. Penman and Reggiani (2018) suggest that the Book-to-Price ratio (B/P) is misleading. Low B/P values reflect uncertainty about future cash-flows rather than "cheap" buying opportunities. Additionally, Altman et al. (2016) argue that, amongst others, profitability is time-varying and find low predictive ability for longer time-periods. Specifically, while they find that profitability ratios like Return on Assets (ROA) provide efficient accuracy for a short horizon of two years , these measures fail to be consistent predictors over a ten-year horizon. They observe that in multivariate models, profitability is rendered largely insignificant when tested against solvency measures, such as the Equity Ratio, which dominates the prediction of distress irrespective of the horizon length.

2. **Predictive Ability:** Traditional bankruptcy models heavily relied on linear combinations of ratios, like the Z-score introduced by (Altman, 1968). Since then, a variety of risk-management models have been introduced, culminating in modern ML applications, such as Random Forest (RF) or support vector machines (Barboza et al., 2017). Even though logit/probit models work reasonably well, Jones et al. (2017) also recommends the use of more advanced machine-learning methods, like AdaBoost or RF.

3. **Misalignment of Prediction Horizons and Objectives:** While traditional bankruptcy models aim to minimize credit risk, they often fail to minimize market risk. A fundamental disconnect exists between *legal insolvency* and *market implosion*.

   - **Lagging Indicators:** Legal bankruptcy is frequently the final stage of a long deterioration process. By the time a traditional Altman Z-Score or structural model flags a company as distressed, the market has often already priced in the failure, resulting in a "Zombie state" where the asset lingers at depressed valuations (Tewari et al., 2024). For an equity investor, the capital is lost at the *implosion* event, not the bankruptcy filing.

- **The Cost of False Positives (Type I Errors):** In the context of equity indexing, the cost of a False Positive is mainly the opportunity cost. As noted in the "Agony and Ecstasy" framework, index returns are driven by a small tail of extreme winners. Traditional models, which penalize negative skewness too aggressively, risk flagging volatile but successful growth stocks as "distressed." Excluding a future "Ecstasy" stock (like NVIDIA) due to a conservative model, which produces too many False Positives, would severely underperform the benchmark, negating the benefits of avoiding the "Agony" stocks.

1.3 Research Gap and Proposed Methodology

1.3.1 Implications of Current Limitations

Ben Jabeur et al. (2021, 2023) have demonstrated that ML models, such as Gradient Boosting and Neural Networks, outperform statistical models in predicting financial distress by capturing non-linear relationships between variables. Tewari et al. (2024) have built on these insights and applied modern ML-techniques to market-based risk-modelling, for which they find that XGBoost can predict up to 61% of implosions in the test set with a false positive rate of less than 3%.

While the literature has identified a variety of ML-methods for credit-risk modelling, there is a lack of research on applying modern ML techniques to the specific problem of *market-based* catastrophic stock declines. Ensemble methods, like bagging and boosting, or Neural Networks are well suited to not only improve predictive accuracy, but also to capture the time-series dynamics of individual stock risk.

1.3.2 Proposed Approach: The "Crash-Filtered" Index

This thesis proposes bridging the gap between distress prediction and active index construction. While Tewari et al. (2024) established the concept of Catastrophic Stock Implosion (CSI), it remains unknown whether these insights can be operationalized into a viable risk-mitigation strategy. To address the "Agony and Ecstasy" dilemma, this research moves from "pure volatility forecasting" to "probabilistic implosion modeling" through the following three-stage methodology:

1. **Model Fitting and Selection:**
   The research will model the probability of market-risk loss, specifically the CSI event, by adopting the definition established by Tewari et al. (2024). The methodology extends current literature by moving beyond static snapshots of financial health. In addition to advanced ensemble methods (e.g., XGBoost, CatBoost) capable of handling non-linear interactions, this approach incorporates autoencoders to decode the noisy financial features.

2. **Index Construction:**
   Based on the model outputs, a "Crash-Filtered" equity index will be constructed. This process involves a systematic re-weighting or exclusion mechanism that penalizes index constituents exhibiting a likelihood of implosion exceeding a calibrated threshold. The primary objective is to penalize identified "Agony" candidates (permanent capital loss or capital loss exceeding a threshold) while retaining the "Ecstasy" winners (usually featuring high volatility and returns). This distinction is critical to isolating the alpha generated purely by tail-risk mitigation, rather than by a generic low-beta factor tilt.

3. **Backtesting and Performance Evaluation:**
   The efficacy of the strategy will be validated through out-of-sample backtesting using a rolling-forward cross-validation framework. The performance of the "Crash-Filtered" index will be benchmarked against:

   - The unfiltered market-weighted benchmark (e.g., Russell 3000) to test for alpha generation.
   - Traditional volatility-weighted portfolios (e.g., Minimum Volatility indices) to test for superior drawdown characteristics.

   Evaluation metrics will prioritize risk-adjusted returns and tail-risk characteristics.

Chapter 2

Research Question

Based on the identified problem that standard metrics fail to distinguish between "recoverable volatility" and "permanent implosion" and that perceived quality-signals can be misleading, the following research question could be explored:

2.1 Main Research Question

To what extent does a 'Crash-Filtered' equity index, constructed using a hybrid of Autoencoder-based feature extraction and Gradient Boosting Ensembles, generate superior risk-adjusted returns compared to traditional volatility-weighted strategies?

2.1.1 Sub-Question 1: Autoencoder

Standard ratios (like P/E) are too noisy, but an Autoencoder can find the "hidden" structure of a failing firm.

Does the integration of latent features derived from Autoencoders significantly improve the predictive accuracy (AUC-ROC) of Gradient Boosting models compared to using raw financial ratios alone?

Hypothesis: Autoencoders will successfully denoise volatile market indicators, allowing the Ensemble model to identify "structural" distress earlier than models relying on raw accounting inputs.

2.1.2 Sub-Question 2: The "False Positive" Advantage of Ensemble Methods

Showing that decision trees (handling the non-linearities) are better than the linear Altman Z-Score at saving the "Ecstasy" stocks.

Do Ensemble methods (e.g., CatBoost, XGBoost) exhibit a statistically lower Type I error rate (False Positives) than linear discriminant models (Altman Z-Score) when classifying high-volatility growth stocks?

Hypothesis: Ensemble models will distinguish between 'good' volatility (growth) and 'bad' volatility (implosion) more effectively than linear models, thereby reducing the exclusion of high-performing winners.

2.1.3 Sub-Question 3: The "Tail-Risk" Advantage

Can a portfolio filtered by this hybrid ML-probability score deliver superior downside protection (lower Maximum Drawdown) than a Minimum Volatility benchmark, without sacrificing the upside participation of a market-cap weighted index?

Hypothesis: The Hybrid-Filtered Index will decouple downside risk from upside potential, generating alpha specifically through the avoidance of Catastrophic Stock Implosions.

Chapter 3

Research Design

1. **Unsupervised Feature Extraction (Autoencoder)**
   To address the high dimensionality of the dataset (over 400 features, including TSFEL time-series aggregations and fundamental ratios like `ff_earn_yld` and `ff_cf_sales`), an Undercomplete Denoising Autoencoder is employed.

   The Autoencoder serves two distinct functions in the predictive pipeline:

   (a) **Dimensionality Reduction:** It compresses the noisy 400-dimensional input vector x into a lower-dimensional latent representation vector z (bottleneck layer). This extracts non-linear structural dependencies between variables (e.g., the interaction between decreasing Cash Flow from Operations and volatile Log Returns) that linear PCA would miss.

   (b) **Anomaly Detection Feature:** Following the future work suggested by Tewari et al. (2024), the model calculates the *Reconstruction Error* ($\mathcal{L} = ||x - \hat{x}||^2$). A high reconstruction error serves as a proxy for "abnormal" market behavior, added as an explicit feature to the supervised Ensemble model.

2. **Defining the target variable (Forward-Looking)**
   Consistent with the supervised learning framework of Tewari et al. (2024), the model is trained to predict the onset of a Catastrophic Stock Implosion (CSI).

   A stock is classified as a positive CSI case if it satisfies the following three conditions:

   (a) **Initial Crash:** The cumulative return drops below a threshold $C = -0.8$ (i.e., an 80% drawdown from the trailing peak).

   (b) **Zombie Period:** The stock enters a stagnation period of duration $T = 78$ weeks (approx. 1.5 years).

   (c) **Non-Recovery:** Throughout the zombie period, the cumulative return never exceeds the recovery ceiling $M = -0.2$ (remaining at least 20% below the peak).

   While Tewari et al. (2024) employ a yearly prediction horizon ($h = 12$ months), this study follows this approach and also tests whether this horizon can be relaxed to $h = 6$ months to enhance the temporal precision required for more frequent index rebalancing. The binary target variable $y_{i,t}$ is defined as:

$$y_{i,t} = \begin{cases} 1 & \text{if stock } i \text{ triggers } C = -0.8 \text{ within } [t, t+h] \text{ and satisfies the zombie criteria} \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

   Sensitivity Analysis and Parameter Relaxation:
   The baseline definition of a 1.5-year ($T = 78$ weeks) zombie period is highly conservative. To assess robustness, this study will perform a sensitivity analysis by relaxing the duration parameter T to **less than 1 year** (e.g., $T = 26$ or $T = 52$ weeks). This tests whether the "Agony" of implosion can be effectively captured with a shorter confirmation window, potentially allowing the model to identify distress signals earlier without significantly increasing the False Positive Rate.

3. **Model Development**

   - **Algorithm:** Ensemble methods (XGBoost, CatBoost) are used to predict the probability of CSI ($y_{i,t} = 1$).

- **Validation of Subquestion 1:** To isolate the contribution of the Autoencoder, an internal comparison will be conducted between:
  - *Model A:* Gradient Boosting on Raw Financial Ratios.
  - *Model B:* Gradient Boosting on Autoencoder Latent Features.
- **Validation of Subquestion 2:** The best-performing Ensemble model is compared against:
  - The Naive Baseline (class prior).
  - Logistic Regression.
  - The **Inverted Altman Z-Score**, treated as a continuous risk ranking to allow for direct AUC-PR comparison.

4. **Index Construction**
   The "Crash-Filtered" index is constructed by filtering the CRISP-universe.

   - **Rebalancing:** The portfolio is rebalanced **anually** or **semi-anually** (dependent on the prediction horizon, e.g. $h = 12$ months) to ensure timely removal of deteriorating assets.
   - **Filtering Mechanism:** For each stock, if the predicted probability $\hat{p}_{i,t} > \theta$, it is excluded.
   - **Calibration:** θ is calibrated on the validation set to maximize the F1-score (harmonic mean of Precision and Recall), balancing the cost of False Negatives (holding an implosion) against False Positives (missing a winner).

5. **Performance Evaluation**
   The strategy is backtested using a rolling-forward cross-validation framework. Performance is benchmarked against the **Market-Weighted Index** (unfiltered) and the **MSCI-USA Minimum Volatility Index**.

Chapter 4

Expected Contribution

1. **Methodological Advancement: A Hybrid Unsupervised-Supervised Framework**
   While Tewari et al. (2024) successfully demonstrated the efficacy of supervised learning (XGBoost) for predicting catastrophic implosions, they explicitly identified unsupervised learning as a "promising avenue for research" to model implosions as anomalies. This thesis directly addresses this call by introducing a **Hybrid Autoencoder-Ensemble architecture**. By using an Autoencoder to denoise high-dimensional financial data and extract latent "distress structures," this research contributes a novel feature engineering pipeline that aims to improve predictive performance in noisy market regimes where standard ratios fail.

2. **Practical Operationalization: From Prediction to Investment Utility**
   Existing literature often focuses on the statistical accuracy of distress models (e.g., AUC, Precision, Recall). However, a statistical prediction does not guarantee economic value. This thesis shifts the focus from "pure prediction" to "portfolio construction." By developing a systematic **"Crash-Filtered" Index**, this research validates whether the statistical signal of a Catastrophic Stock Implosion (CSI) can be monetized. It provides empirical evidence on whether avoiding specific "Agony" events generates superior risk-adjusted returns compared to generic Minimum Volatility strategies, thus bridging the gap between machine learning metrics and investor outcomes.

3. **Theoretical Distinction: Disentangling Volatility from Implosion**
   Traditional risk models treat volatility as a proxy for risk, often penalizing high-growth "Ecstasy" stocks that exhibit "good" volatility. This thesis contributes to the "Agony and Ecstasy" framework (Cembalest, 2014) by empirically testing whether "Catastrophic Implosion" is a distinct risk factor separate from standard price variance. By calibrating the model to minimize Type I errors (False Positives), this research seeks to demonstrate that it is possible to decouple downside tail-risk protection from the upside participation of growth stocks—a distinction that standard low-beta factors fail to achieve.

Bibliography

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

Edward I. Altman, Małgorzata Iwanicz-Drozdowska, Erkki K. Laitinen, and Arto Suvas. Financial and non-financial variables as long-horizon predictors of bankruptcy. *Journal of Credit Risk*, 12(4):49–78, 2016.

Flavio Barboza, Herbert Kimura, and Edward I. Altman. Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83:405–417, 2017.

Sami Ben Jabeur, Cheima Gharib, Salma Mefteh-Wali, and Wissal Ben Arfid. Catboost model and artificial intelligence techniques for corporate failure prediction. *Technological Forecasting and Social Change*, 166:120658, 2021.

Sami Ben Jabeur, Nicolae Stef, and Pedro Carmona. Bankruptcy prediction using the XGBoost algorithm and variable importance feature engineering. *Computational Economics*, 61:715–741, 2023.

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.

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, 2024.

Stewart Jones, David Johnstone, and Roy Wilson. Predicting corporate bankruptcy: An evaluation of alternative statistical frameworks. *Journal of Business Finance & Accounting*, 44(1–2):3–34, 2017.

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.