Master Thesis - Research Proposal

Tristan Leiter

December 2, 2025

Abstract

This report demonstrates the integration of R code and its output within a LaTeX document using Sweave. It covers the basic structure of a report, including a summary, chapters with subchapters, and a bibliography.

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 (atleast 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 reasonable 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 for applying modern ML techniquers 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, it remains unknown whether these insights can be operationalized into a viable risk-mitigation strategy.

To address the "Agony and Ecstasy" dilemma, this research aims to move from "pure volatility forecasting" to "probabilistic implosion modeling." The proposed methodology improves upon the status quo in three specific ways:

1. **Advanced Ensemble Modeling:** Utilizing Bagging and Boosting algorithms (e.g., XGBoost, CatBoost) as well as Neural Networks to handle the non-linear feature interactions of raw market and fundamental data.

2. **Temporal Dynamics with Deep Learning:** Extending the work of Ben Jabeur et al. (2023) by incorporating Long Short-Term Memory (LSTM) networks. This allows the model to treat company health as a time-series problem, detecting the *rate of change* in fundamentals that precedes a crash, potentially reducing the "lag" associated with traditional models.

3. **Predictive Exclusion Framework:** Constructing a "Crash-Filtered" index using Rolling-Forward Cross-Validation. Instead of the traditional binary exclusion (listing failure), this framework applies a probabilistic threshold. The goal is to systematically exclude identified "implosion" candidates while retaining the "Ecstasy" winners, thereby isolating the alpha generated purely by risk mitigation.

4. **Performance Evaluation:** The performance of the strategy will be evaluated by comparing the risk-adjusted returns and risk-characteristics of the filtered index against the unfiltered market-weighted benchmark and traditional volatility-weighted portfolios.

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

Does a ML-risk filtered equity index, constructed by excluding stocks with a high implied probability of catastrophic implosion, generate statistically significant superior risk-adjusted returns compared to the market-weighted benchmark and traditional volatility-filtered indices?

2.1.1 Sub-Question 1: Predictive Superiority

Can Machine Learning models (e.g., CatBoost, XGBoost or Neural Networks) calibrated to the definition of "Catastrophic Implosion" outperform traditional distress models (e.g., Altman Z-Score) in distinguishing between permanent capital loss and temporary drawdown?

This addresses the "Zombie State" gap, testing if ML can identify stocks that implode but do not necessarily go bankrupt. It also leverages the finding that advanced algorithms like CatBoost, XGBoost and Neural Networks provide higher accuracy in distress prediction than traditional linear models.

2.1.2 Sub-Question 2: Portfolio Distinction

Does the "Implosion-Filtered" portfolio exhibit lower risk characteristics (measured by maximum drawdown and expected shortfall) than standard risk-controlling models (low volatility, low beta, ...)?

This tests the hypothesis that avoiding *implosions* (permanent loss) provides a distinct risk profile compared to avoiding *volatility* (temporary noise), potentially allowing the investor to retain exposure to high-growth, high-volatility winners that do not implode.

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.