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

The theoretical foundation of modern portfolio management relies on the diversification of idiosyncratic risk. However, empirical evidence suggests that the distribution of individual stock returns is not normally distributed but is instead characterized by extreme skewness and "fat tails".

Research by J.P. Morgan Asset Management highlights a 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. 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 non-linear dynamics preceding permanent capital decline.

1.2 Status Quo

Traditional risk management frameworks typically rely on linear risk metrics, such as realized volatility or market Beta, to filter investment universes. The prevailing assumption is that "Low Volatility" equates to safety. However, this approach is insufficient for avoiding catastrophic risk for three primary reasons:

1. **The Quality Trap:** Perceived fundamental health is not a shield against catastrophe. As noted in the J.P. Morgan analysis, over half of catastrophic decliners were profitable at their peak (Cembalest, 2014, 2024). Incorporate profitability. For example, standard metrics, such as Book-to-Price (B/P) can be misleading, resulting in a "Value-Trap" (Penman and Reggiani, 2018).

2. **Linearity:** Additionally, standard linear filters fail to flag these "healthy" stocks before they break. Furthermore, traditional bankruptcy models, such as the Altman Z-Score, heavily rely on linear combinations of ratios which may be outdated or insufficient for capturing modern market dynamics or are implemented via generalized linear models (logistic-regressions), which are shown to have lower predictive power than more advanced algorithms such as CatBoost or XGBoost.

3. **The "Zombie State" Gap:** Bankruptcy prediction models can be potentially valuable, but are prone to produce too many false positives or lag the market reaction. Many stocks enter a "zombie state" where, after an implosion, they linger in the market with depressed valuations without legally declaring bankruptcy. Investors are thus left holding illiquid assets that traditional distress models fail to flag until the capital is already lost. In addition, traditional models focus on negative skewness. It can be useful for short-term investing, but is usually to sensitive for long-term holdings as they flag stocks that suffer from temporary distresses, but which recover quickly. Thus, it would produce too many false-positives on the long-run.

1.2.1 What is unknown?

While the literature on bankruptcy prediction is extensive, there is a lack of research applying modern Machine Learning (ML) techniques to the specific problem of *market-based* catastrophic declines in the cross-section of prior "healthy" index constituents.

Recent work has demonstrated that ML models, such as Gradient Boosting and Neural Networks, outperform traditional statistical models in predicting financial distress by capturing non-linear relationships between variables (Ben Jabeur et al., 2021, 2023), (Tewari et al., 2024). For instance, models like CatBoost and XGBoost have shown superior classification accuracy in failure prediction compared to discriminant analysis and logistic regression.

These results by Tewari et al. (2024) could be built upon by further incorporating LSTM for time-series forecasting. In addition, it remains unknown whether their insights can be useful as a viable risk-mitigation strategy. For example, one could create a "Crash-Filtered" index constructed by systematically excluding these identified "implosion" candidates and explore whether improvements in risk-adjusted returns can be delivered compared to traditional risk-mitigation techniques.

1.2.2 What could be improved?

The current paradigm of index construction includes all constituents unless they fail listing requirements or suffer massive market cap erosion (a reactive approach). By shifting to a *predictive exclusion* framework using Machine Learning, there is potential to mitigate the "Agony" side of the return distribution. This involves moving beyond "pure volatility forecasting" to "probabilistic implosion modeling," potentially leveraging a set of raw-data features without computing any ratios (these dependencies can be picked up efficiently via the employed algorithms).

Chapter 2

Research Questio

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

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.

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.