Planting Portfolios

Panel Trees in the Swedish Equity Market

Erik Jin
Michael Brusis

Bachelor Thesis
Stockholm School of Economics
2025

Abstract:

This study evaluates the performance of the P-Tree algorithm (Cong et al., 2025) in the Swedish equity market, a setting characterized by lower liquidity, fewer listed stocks, and limited historical data. The P-Tree algorithm identifies return patterns by splitting stocks based on firm characteristics, capturing both linear and non-linear relationships.

In the Swedish context, the in-sample Sharpe ratio of the P-Tree portfolio was 0.92, exceeding the market portfolio's 0.64. Out-of-sample, the Sharpe ratio was 0.53 when training on the first half of the sample and testing on the second half, and 0.48 in the reverse scenario. The P-Tree strategy produced a statistically significant Fama–French three-factor α of 9.61% in-sample, and 3.21% out-of-sample when predicting from the past to the future. In the reverse future-to-past prediction, the α was 3.39%, which was not statistically significant at the 10% confidence level.

The P-Tree model offers flexible structure, with the number of splits adapted to the available data. In markets with limited historical or cross-sectional information, a shallow-tree configuration can be implemented that focuses on the most informative characteristics, avoiding overfitting. In data-rich environments, a deep-tree configuration can capture complex, non-linear relationships while maintaining robustness. This adaptability ensures that P-Trees balance predictive power with data availability, delivering interpretable and reliable predictions. Such properties make them a promising tool for portfolio optimization, risk management, and factor-based investing in the finance industry.

Keywords:

Machine learning, portfolio optimization, decision tree, efficient frontier, Sharpe ratio

Authors:

Erik Jin (26057)
Michael Brusis (25895)

Tutor:

Maíra Sontag González, Ph.D., Department of Finance

Examiner:

Ramin P Baghai, Professor, Department of Finance

Bachelor Thesis in Finance

Bachelor Program in Business and Economics Stockholm School of Economics

1 Introduction

It has long been understood that you should not put all your eggs in one basket. This metaphor extends naturally to investing: rather than allocating all your capital to a single stock, you can reduce risk by diversifying across multiple assets. This idea forms the foundation of Modern Portfolio Theory (MPT), introduced by economist Markowitz (1952). The objective of MPT is to maximize expected return for a given level of risk, or equivalently, minimize risk for a given expected return. By plotting expected return against risk (measured by the standard deviation of portfolio returns) for all possible portfolios, one obtains the efficient frontier, which is the set of portfolios offering the optimal trade-off between risk and return.

When a risk-free asset is introduced, the line tangent to the efficient frontier defines the Capital Market Line (CML). The point of tangency represents the optimal risky portfolio (the tangency portfolio), which, under the assumptions of the Capital Asset Pricing Model (CAPM), corresponds to the market portfolio. Portfolios along the CML dominate all others, offering the highest expected return for a given level of risk. The excess return over the risk-free rate per unit of risk is measured by the Sharpe ratio, a key metric in portfolio construction.

Although Modern Portfolio Theory provides a theoretical framework for constructing efficient portfolios, its practical implementation depends on accurate estimates of expected returns, variances, and covariances. As the number of assets increases, the number of parameters that must be estimated grows quadratically, making reliable estimation increasingly challenging. To address these challenges, empirical asset pricing research has developed alternative, data-driven methods for portfolio formation. One of the most influential approaches is the construction of characteristic-sorted portfolios, where stocks are grouped based on observable firm characteristics such as size, value, or momentum. Characteristic-sorted portfolios are designed to capture how expected returns vary with firm characteristics. Stocks are ranked based on an observable variable such as size, book-to-market ratio, or momentum, and grouped into quantiles, with deciles or quintiles being most common in the literature. This allows researchers to examine how average returns differ across these groups. The difference in returns between high- and low-characteristic portfolios captures the return premium associated with a given characteristic. In the finance literature, characteristic-sorted portfolios are commonly used as benchmarks for evaluating portfolio performance. Their appeal lies in their simplicity and transparency, which makes them easy to interpret and replicate. Moreover, they are empirically grounded, as numerous studies have documented that systematic return patterns are linked to well-known factors such as size, value, and momentum. As a result, characteristic-sorted portfolios not only capture meaningful risk factors but also exhibit stability and robustness, since the underlying stock characteristics tend to remain relatively consistent over time.

In recent years, machine learning techniques have been increasingly applied to portfolio optimization. Unlike traditional methods, which often rely on linear assumptions and parametric estimates, machine learning algorithms can capture complex, non-linear relationships between asset returns and risk factors. These models are particularly well-suited for high-dimensional datasets, allowing the integration of multiple sources of information, including firm fundamentals and macroeconomic factors. Additionally, once trained, these models can scale to large asset universes. One novel approach is the P-Tree (Panel Tree) algorithm introduced by Cong et al. (2025), which extends traditional tree-based methods to panel data structures with the splitting criteria chosen to maximize an objective function. In this context, portfolio optimization begins with the entire asset universe at a root node. The algorithm then systematically splits assets into child nodes based on firm characteristics

(such as trading volume or book-to-market ratios), forming distinct portfolios at each leaf node. The algorithm is covered in more detail in section 3.2.

Unlike standard decision trees that optimize prediction accuracy locally, the P-Tree model uses a global optimization criterion of maximizing the Sharpe ratio of the mean-variance efficient portfolio constructed from all leaf portfolios. This approach maintains a time-invariant tree structure, ensuring consistency across periods while capturing asymmetric interactions between characteristics. For example, liquidity measures may predict returns differently for high-earnings versus low-earnings stocks. By iteratively splitting the cross-section and re-estimating the tangency portfolio, P-Tree generates diversified test assets that span a more efficient frontier than conventional sorted portfolios, while remaining interpretable through its graphical tree structure.

However, machine learning approaches typically excel when there is high-dimensional data with complex interactions, and they require large amounts of data to train. As such, they have been predominantly developed and tested in highly liquid markets such as the US stock market, where extensive information about individual stocks is readily available. This raises an important question: do these sophisticated models maintain their strong performance in less liquid markets with sparser information and fewer observable characteristics? Testing machine learning methods in such environments provides critical insights into their robustness and external validity beyond the well-documented, data-rich settings where they were originally designed. Thus, our research question is: **Can the panel-tree algorithm still be used for portfolio construction in the Swedish stock market?**

2 Literature Review

2.1 Existing Literature

Machine learning methods have demonstrated substantial improvements in cross-sectional return prediction for US equity markets. Gu et al. (2020) show that tree-based models and neural networks deliver high Sharpe ratios and significant alphas relative to traditional factor models. Crucially, they emphasize that shallow architectures with strong regularization outperform deep models because they avoid overfitting when predictive signals are weak relative to noise. Their best-performing strategies are market-neutral long-short portfolios with low market beta but high risk-adjusted returns. Bryzgalova et al. (2023) extend these findings by applying gradient-boosted decision trees to portfolio construction. They demonstrate that single-split stumps capture the majority of economically meaningful predictability, while additional splits primarily fit noise rather than genuine return patterns. This emphasis on simple models reflects a broader recognition that weak return predictability favors interpretable specifications over complex alternatives.

However, theoretical and empirical evidence suggests that machine learning performance depends on market size and data quality. Martin and Nagel (2021) formalize the dimensionality problem in asset pricing: when characteristics outnumber observations, even optimal investors must use shrinkage or sparsity to avoid overfitting. They show that unconstrained complex models generate misleading in-sample predictability, while true out-of-sample performance is best when models are sparse. High in-sample significance often reflects estimation error rather than exploitable patterns. Ghandar et al. (2016) provide supporting evidence, demonstrating that complex prediction models underperform simple alternatives in dynamic financial markets. Freyberger et al. (2024) recommend minimum leaf size constraints and early stopping to prevent decision trees from fitting noise,

particularly in small panels.

Most directly relevant to smaller markets, Didisheim et al. (2024) demonstrate that complex machine learning models systematically overfit in international equity markets outside the United States. They attribute this underperformance to limited cross-sectional coverage and weaker predictive signals. They recommend conservative specifications, such as shallow trees and ridge regression for non-US markets. Deeper architectures provide no economic value over conservative specifications in data-constrained settings and serve only to add noise. This finding is reinforced by evidence that factor premiums themselves are country-specific. Griffin (2002) establishes that asset pricing patterns documented in US data do not automatically generalize internationally, with local factor implementations explaining returns better than global factors. Ammann and Steiner (2023) confirm this pattern specifically for Nordic markets, finding that factor premiums across Denmark, Finland, Norway, and Sweden are weaker and less stable than in US data, attributed to smaller market size, lower liquidity, and higher idiosyncratic risk. Nevertheless, Drobetz and Otto (2021) show that machine learning can improve return forecasting in European markets when models are adequately trained and tuned to avoid overfitting, suggesting that careful implementation matters as much as market size.

This thesis extends Cong et al. (2025), who develop the Panel Tree (P-Tree) framework for portfolio construction in the US stock market. P-Trees belong to the broader literature on latent factor models in asset pricing, which includes deep learning models for nonlinear factor modeling and various PCA-based approaches. While these methods can capture variable nonlinearity and asymmetric interactions, P-Trees offer a distinct advantage with graphical interpretability combined with economic guidance. P-Trees recursively split the cross-section of stocks to maximize the Sharpe ratio of portfolios formed at each leaf node, maintaining a time-invariant structure that facilitates interpretation. The algorithm incorporates regularization through minimum leaf size constraints and employs a global optimization criterion rather than local prediction accuracy. Cong et al. (2025) demonstrate that P-Trees achieved strong out-of-sample performance in US data, typically producing up to 9 splits, with their primary single tree, that identify complex interactions between characteristics. This raises the question of whether P-Trees maintain their advantages in smaller markets with limited data and fewer available characteristics.

Despite data limitations, several firm characteristics exhibit predictive power in smaller international markets. Soliman (2008) decomposes return on equity and finds that asset turnover contains information about future returns beyond simple profitability ratios, as it is positively correlated with subsequent earnings changes and future stock returns even after controlling for other ROE components. Hou et al. (2022) confirm that asset turnover remains a robust predictor across multiple international samples, including European markets, and is economically significant even in smaller cross-sections. Similarly, cash holdings appear linked to return premiums in constrained markets. Simutin (2010) documents that firms with excess cash earn higher subsequent returns than firms with low or negative excess cash, even after controlling for standard risk factors. Bates et al. (2023) extend this finding internationally, showing that high cash-to-asset ratios command a return premium in smaller European markets, with the strongest effect in countries with higher idiosyncratic volatility and weaker corporate governance. Beyond individual characteristics, the market structure itself may favor certain strategies. Bali et al. (2024) find that the premium for market-neutral strategies is highest in smaller, less liquid markets due to higher idiosyncratic volatility and limited arbitrage capital. This suggests that a market-neutral P-Tree approach may be particularly well-suited to the Swedish context despite data constraints.

2.2 Contribution to Literature

Despite extensive research on machine learning in US asset pricing, P-Trees remain untested in smaller international markets. This thesis provides the first application of the P-Tree methodology outside the United States, examining whether interpretable machine learning methods maintain their advantages in data-constrained environments. The Swedish equity market represents a critical test case as a developed market with fewer listed firms than the United States, allowing us to assess how sample size affects machine learning performance in a high-quality institutional setting. To enable this analysis, we developed a custom data processing pipeline integrating monthly market data from Swedish House of Finance (2025b) with annual accounting data from Swedish House of Finance (2025c), applying forward-filling logic to ensure data availability while preventing look-ahead bias. We train single P-Tree models to construct optimized portfolios and evaluate their performance against standard Swedish asset pricing factors, testing whether P-Trees can extract risk premiums beyond those captured by traditional factor models.

3 Methodology

3.1 Data Gathering and Preparation

Obtaining a clean, quality dataset is arguably one of the most challenging and time-consuming parts of training any machine learning algorithm. P-Trees are no exception to this challenge. The problem becomes even more severe in smaller markets with limited coverage, liquidity, and fewer publicly traded equities. To address these constraints, a custom data preparation pipeline was built that combined several available datasets into one cleaned final dataset suitable for training.

The choice of data sources is critical for the validity of any empirical study. Unlike the US market, where widely used databases such as CRSP exist, there is no single database for the Swedish market that contains all firm-level fundamentals and market variables required for this study. We evaluated several potential data sources, including "LSEG" (2025), but for our specific sample and variable requirements its coverage was incomplete for some companies and accounting items. This led us to focus on databases provided by the Swedish House of Finance. The FinBas dataset from Swedish House of Finance (2025b) provides the universe of listed companies on the Stockholm Stock Exchange together with monthly market data and selected firm characteristics. The Serrano dataset from Swedish House of Finance (2025c) complements this with detailed financial statement data for Swedish firms. Comparable information may be available in other commercial databases such as Bloomberg, but these were not accessible within the scope of this research.

The data preprocessing consisted of multiple steps. Data points outside the time frame were filtered out. Missing values were handled. Duplicate variables present in multiple datasets were removed. Unnecessary variables were excluded from the model. As previously mentioned, the time window chosen was from June 1998 to November 2019. This is due to the availability of the Fama-French factors provided by the Swedish House of Finance which only had data up to 2020. The lower bound is determined by both the availability of the Serrano database, with the earliest observation in 1997, and data quality constraints with limited observations and insufficient coverage for many firm characteristics for a reliable cross-sectional analysis. Part of the filtering included removing multiple share types from the same firm. If multiple share classes were included, the firm-level characteristics used as explanatory variables are identical across share classes, while the market variables (prices, returns, liquidity) differ. This creates multiple observations with the same

explanatory features but different outcomes. In a prediction or regression setting, it can bias the model. Therefore, the most liquid share class was chosen to represent the firm. The liquidity was measured primarily with turnover in SEK, but traded volume was used as a secondary tie-breaker. Another part of the process was to filter out foreign stocks. These stocks are defined as stocks of firms based in other countries. The FinBas database included stocks belonging to Finnish firms which were not part of the analysis and thus filtered out. Trading data from four Stockholm based Swedish equity markets were included. These were ranked according to exchange quality to ensure consistent price information and avoid duplicating observations for stocks listed on multiple venues. Nasdaq Stockholm (Main Market), the primary regulated exchange in Sweden, was given the highest priority (Rank 1). The secondary or alternative markets were ranked as follows: Nasdaq First North (Growth Market, Rank 2), Nordic Growth Market (NGM, Rank 3), and Spotlight / AktieTorget (Rank 4). This ranking system ensures that when a stock is listed on multiple exchanges simultaneously, only the highest-priority venue is used in the dataset.

Only variables that could be used to replicate the variables used by Cong et al. (2025) were kept. This was for the variables from both the FinBas and Serrano datasets. Variables that had low coverage were also excluded, even if they were used by Cong et al. (2025). For example, dividend yield only had a coverage of 0.44%. With such low coverage the variable would not have any significant predictive power.

The next step was to merge the data. Before merging the cleaned FinBas and Serrano datasets, a mapping table was created. The FinBas dataset used ISIN as identifier while Serrano used organization number (org nr) as identifier. The "LSEG" (2025) database was used to create this mapping, as it included both ISIN and tax-id. In Sweden, the tax-id is constructed from the organization number by adding the string "SE" to the start and the integers "01" to the end. Thus, to retrieve the organization number, these were stripped off. However, LSEG coverage was incomplete. Approximately half of the stocks in our sample could not be mapped using LSEG data and had to be matched manually by cross-referencing company names and identifiers from multiple sources. The final mapping table was constructed as a dictionary with ISINs as keys and organization numbers as values.

One problem that needed to be accounted for when merging the two datasets was that FinBas had monthly data while Serrano only had annual data. They could therefore not be merged directly. One option for merging the data was to convert the FinBas data from monthly to annual data by computing the average for the monthly data points across every calendar year. The main concern with this approach would be the loss of granularity in the dataset, which is of particular concern since this would reduce the number of time points from 258 to only 22. This issue is further amplified because most stocks did not have full coverage across the whole time window studied (see Section 4).

The method we used was to merge the datasets by forward-filling annual accounting data across multiple months. The main drawback with this approach is that the model loses some predictive power since the same accounting variables are used to predict different monthly returns. However, the motivation for proceeding with this approach is that this information structure reflects real-world conditions. An investor's knowledge of the most recent accounting information for a firm is based on the firm's most recent annual report. After the tables were merged, new characteristics that use information from both tables were created. These included return on assets, book-to-market, earnings-to-price ratio, sales-to-price, and cashflow-to-price. The final filter removed all rows containing more than four missing values. This step was implemented to reduce noise in the dataset and improve the signal-to-noise ratio for subsequent analysis.

The final step involves lagging the explanatory variables to prevent look-ahead bias. In practice, only past information is available when making investment decisions. Using contemporaneous information to predict stock prices would artificially inflate predictive power and fail to reflect realistic trading conditions. By lagging the variables, we ensure that the model relies solely on information observable at the time of prediction. The number of lags applied to each variable depends on its category, as shown in Table 2. As a consequence of using lag, the total number of usable observations in the dataset is reduced. Table 1 shows the full data filtering pipeline and the number of remaining observations after each step.

Table 1: The data filtering pipeline applied to construct the dataset. Steps include: removing non-Swedish equities, excluding over-the-counter or off-exchange stocks, keeping only one exchange per stock, including only stocks with annual reports from the Serrano database, and filtering out observations with more than four missing values. The filtration reduces the number of observations from approximately 112,000 to 39,524.

Step	Records	Description
1. Raw FinBas monthly data	∼112,000	All FinBas records
2. Filter SE/SEK stocks only	∼105,000	Swedish stocks in SEK
3. Remove OTC/off-exchange	∼95,000	Liquid exchange-traded only
4. De-duplicate by ISIN-date	∼95,000	One record per firm-date
5. Merge with Serrano (accounting)	∼67,000	Inner join via ISIN-OrgNr
6. Apply lags and filters	∼61,000	Remove missing targets
7. Final dataset	39,524	Ready for P-Tree

Table 2: Variable categories and applied lags to prevent look-ahead bias. Accounting and growth variables have a 6-month publication lag, market variables 1 month, momentum 0–12 months depending on the window, CHPM 6 months, and CHCSHO 12 months.

Category	**Variable**	**Lag / Methodology**
Accounting	Balance Sheet & Income Statement	6-month publication lag data at t corresponds to FYE $\leq t-6$
Market	Market Cap (ratios)	Current month t (e.g., B/M)
	Returns (Target)	Lead return $t+1$
Momentum	MOM1M	Return in t
	MOM6M	Cum. return $t-6$ to $t-1$
	MOM12M	Cum. return $t-12$ to $t-1$
	MOM36M	Cum. return $t-36$ to $t-12$
	MOM60M	Cum. return $t-60$ to $t-12$
	SEAS1A	Return in $t-12$
Growth	AGR	Total Assets YoY $t-18$ to $t-6$
	SGR	Sales YoY $t-18$ to $t-6$
	LGR	LT Debt YoY $t-18$ to $t-6$
Changes	CHCSHO	Shares outstanding YoY $t-12$
	CHPM	Profit Margin change $t-18$ to $t-6$

3.2 P-Tree Algorithm

This thesis employs the P-Tree algorithm developed by Cong et al. (2025) to construct portfolios from an asset universe consisting of equities available in the Stockholm stock exchange. The final portfolios are the leaf nodes, which are constructed such that they can fulfill the global objective which is to construct the efficient frontier. The algorithm, provided by the authors on GitHub Cong et al. (2025), was implemented in R. Prior to modeling, the data were pre-processed in Python using the `pandas` library and then imported into R for use in the P-tree algorithm.

The P-Tree algorithm works on panel data with n assets and T time periods. Assuming a total of j splits, there will be $j+1$ leaf basis portfolios. The returns of these portfolios form a matrix

$$\mathbf{R}_t^{(j)} = [R_{1,t}^{(j)}, ..., R_{n,t}^{(j)}, ..., R_{j+1,t}^{(j)}] \in \mathbb{R}^{T\times(j+1)}$$

where $R_{n,t}^{j}$ represents a length T vector containing the excess returns over the T periods of the n:th leaf node. The latent factors $f_t^{(j)} \in \mathbb{R}^T$ is a time series vector that represents the weighted return of the portfolios after splitting the asset universe into sub-portfolio based on characteristics

$$f_t^{(j)} = \mathbf{R}_t^{(j)} w^{(j)}$$

where $w^j \in \mathbb{R}^{(j+1)}$ is a vector of the weights for the leaf basis portfolios. The optimal weights are calculated with

$$\mathbf{w}^{(j)} \propto \left(\hat{\mathbf{E}}[\mathbf{R}_t^{(j)}\mathbf{R}_t^{(j)'}] + \gamma\mathbf{I}\right)^{-1} \hat{\mathbf{E}}[\mathbf{R}_t^{(j)}] \tag{1}$$

and where each row sums to 1 to ensure that the total weight of all leaf basis portfolios is 1. $\hat{E}[\mathbf{R}_t^j] \in \mathbb{R}^{j+1}$ and $\hat{E}[\mathbf{R}_t^j\mathbf{R}_t^{j'}] \in \mathbb{R}^{(j+1)\times(j+1)}$ are the sample mean excess return and the second moment matrix for the leaf basis portfolios. This is a variation of optimal weight solution derived from Markowitz's mean-variance optimization framework

$$\begin{aligned} &\min \mathbf{w}'\hat{\Sigma}\mathbf{w} \\ &\text{subject to } \mathbf{1}'\mathbf{w} = 1 \\ &\text{subject to } \mathbf{w}'\hat{\mu} = \mu \end{aligned}$$

where $\mathbf{w}$ are the optimal weights, $\hat{\Sigma}$ is the covariance matrix, $\hat{\mu}$ is the expected return and μ is the target return. This is a constrained optimization problem that can be solved using the method of Lagrange multipliers. The weight constraint can be accounted for by first lifting the constraint, and later rescaling the optimal weights such that they sum to 1.

$$\begin{aligned} \mathcal{L}(\mathbf{w},\lambda) &= \frac{1}{2}\mathbf{w}'\hat{\Sigma}\mathbf{w} - \lambda(\mathbf{w}'\hat{\mu} - \mu) \\ \frac{\partial\mathcal{L}}{\partial\mathbf{w}} &= \hat{\Sigma}\mathbf{w} - \lambda\hat{\mu} = 0 \\ \frac{\partial\mathcal{L}}{\partial\lambda} &= \mathbf{w}'\hat{\mu} - \mu = 0 \end{aligned} \tag{2}$$

Let the optimal weight be defined as w^*. The first partial derivative gives $w^* = \lambda\hat{\Sigma}^{-1}\hat{\mu}$ which sets the weights proportional to values that maximizes the Sharpe ratio $\hat{\Sigma}^{-1}\hat{\mu}$, and the using the second partial derivative we obtain

$$w^* = \frac{\mu}{\hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}}\hat{\Sigma}^{-1}\hat{\mu}$$

which gives the weights that optimizes the Sharpe ratio given a level of return μ. The weights can be rescaled $\mathbf{w}^* = \frac{\mathbf{w}}{\mathbf{1}'\mathbf{w}}$ such that they sum to 1.

The algorithm provided by Cong et al. (2025) is a variation of this methodology. They use an unconstrained version that only tries to find weights that optimizes the Sharpe ratio and there is no target return constraint. Instead of using the covariance matrix $\hat{\Sigma}_F$ they use the second moment matrix $\hat{E}[\mathbf{R}_t^{(j)}\mathbf{R}_t^{(j)'}]$ This is a good approximation if the mean matrix defined as $\hat{\mu}_F\hat{\mu}'_F$ has small values in comparison to $\hat{\Sigma}_F$

$$\hat{E}[\mathbf{R}_t^{(j)}\mathbf{R}_t^{(j)'}] = \hat{\Sigma}_F + \hat{\mu}_F\hat{\mu}'_F \approx \hat{\Sigma}_F, \text{ if } \hat{\Sigma}_F \gg \hat{\mu}_F\hat{\mu}'_F$$

This can be justified as the returns often have small mean in comparison to volatility. Their algorithm also includes a shift in the matrix inversion of the size $\gamma\mathbf{I}$. γ is a parameter which stabilizes the matrix inversion which can be a problem if the matrix is ill-conditioned. This can happen if the returns of the leaf basis portfolios are close to or completely linearly dependent. By including this shift, the matrix $\hat{\mathbf{E}}[\mathbf{R}_t^{(j)}\mathbf{R}_t^{(j)'}] + \gamma\mathbf{I}$ becomes better conditioned. This is a trade-off between computational stability and reliability and optimal weight selection.

The algorithm begins with an initialization stage in which all stocks are grouped into a single root node representing the market portfolio, $\mathbf{R}_t^{(0)} \in \mathbb{R}^T$. This is the value weighted return of the asset universe across the T periods. Before the first split, firm characteristics which are the dependent variables in the model, are normalized cross-sectionally to the range $[-1, 1]$ within each period. For the first split, the model evaluates splits from the pre-specified thresholds $c_m \in [-0.6, -0.2, 0.2, 0.6]$ for each characteristic. This splits the assets in the parent node into quintiles based on their value for a specific characteristic. For the first split, the parent node includes all the assets. The algorithm tests splits $\tilde{c}_{k,m}$ where k denotes the characteristic. The value of characteristic k for an asset is denoted z_k and whether the asset is placed in the left child node depends on whether $z_k \leq \tilde{c}_{(k,m)}$ where $\tilde{c}_{(k,m)}$ denotes the current threshold for characteristic k. If it is not, the asset is placed on the right child node. The algorithm chooses the one threshold for the characteristic that maximizes the Sharpe ratio

$$\mathcal{L}(\tilde{c}_{k,m}) = \sqrt{\hat{\mu_F}'\hat{\Sigma}_F^{-1}\hat{\mu}_F} \tag{3}$$

where $\hat{\Sigma}_F^{-1}\hat{\mu}_F$ are the optimal weights derived from Markowitz's unconstrained mean-variance framework, see equation 2. For the first split, $F = f_t^{(1)}$, where $f_t^{(1)}$ is a scalar time series of the return of the sum of the weighted portfolios $\mathbf{R}_t^{(1)} = [R_{1,t}^{(1)}, R_{2,t}^{(1)}]$, where $R_{1,t}^{(1)}$ represents the return of the assets below the threshold and $R_{2,t}^{(1)}$ represents the assets above the threshold. The return of the first factor is calculated by choosing optimal weights for the two sub-portfolios based on a characteristic and a threshold, and taking the weighted sum of their respective returns.

The economic interpretation can be better understood with the following example. Let the split be for the characteristic return on asset (ROA) (full list of abbreviations in Appendix 7), and let the optimal threshold derived from iterating through the pre-specified thresholds in c_m, splitting the assets based on the threshold, and choosing the one that maximizes the Sharpe ratio from equation 3, given the optimal weights given by equation 1. If the threshold was calculated to be 0.2 and the optimal weights for the portfolios are calculated to be $w_1^* = 0.3, w_2^* = 0.7$, then the economic interpretation would be that the assets with return on asset values in the top 40 percentiles have

higher returns, and that the optimal risk adjusted return can be achieved by constructing a portfolio with 30% of total capital allocated to assets in the bottom 60 percentiles of return on asset (ROA), and 70% capital allocated to asset in the top 40 percentiles of return on asset. The factor for the first split is calculated as

$$f_t^{(1)} = w_1^* R_{1,t}^{(1)} + w_2^* R_{2,t}^{(1)}$$

and the sample mean $\hat{\mu}_F$ which represents the mean excess return achieved from the specific split

$$\hat{\mu}_F = \mathrm{E}[f_t^{(1)}]$$

with the covariance matrix $\hat{\Sigma}_F = \mathrm{Var}[f_t^{(1)}] \in \mathbb{R}$ for the first step.

For the second step, one of the child nodes generated from the first step becomes an internal node and is split. After the second split there will be 3 portfolios, $\mathbf{R}_t^{(2)} = [R_{1,t}^{(2)}, R_{2,t}^{(2)}, R_{3,t}^{(2)}]$. The split is chosen to maximize the Sharpe ratio in equation 3 using the weights from equation 1 with the new $\mathbf{R}_t^{(2)} = [R_{1,t}^{(2)}, R_{2,t}^{(2)}, R_{3,t}^{(2)}]$. The optimal weights are $\mathbf{w}^* = [w_1^*, w_2^*, w_3^*]'$ and the new factor $f_t^{(2)}$ is calculated as

$$f_t^{(2)} = \mathbf{R}_t^{(2)} \mathbf{w}^* = \sum_{i=1}^{3} w_i^* R_t^{(i)}$$

with the new sample mean vector

$$\hat{\mu}_F = \begin{bmatrix} \mathrm{E}[f_t^{(1)}] \\ \mathrm{E}[f_t^{(2)}] \end{bmatrix}$$

and the new covariance matrix

$$\hat{\Sigma}_F = \begin{bmatrix} \mathrm{Var}[f_t^{(1)}] & \mathrm{Cov}[f_t^{(1)}, f_t^{(2)}] \\ \mathrm{Cov}[f_t^{(1)}, f_t^{(2)}] & \mathrm{Var}[f_t^{(2)}] \end{bmatrix}$$

For the third step and onward, the same procedure is applied: For each leaf node, calculate the Sharpe ratio for splitting that node given a characteristic and a threshold. Choose to split the node for a given threshold and characteristic that maximizes the Sharpe ratio. The algorithm is terminated when either the maximum depth is reached, the minimum leaf size defined by the minimum number of assets in a leaf, is reached, or that any further split does not increase the Sharpe ratio. The hyperparameters will be covered in section 3.4. Being a greedy algorithm, the sequential growth of the tree ensures computational feasibility compared to a full enumeration of possible sortings Cong et al. (2025). The final output of the algorithm after J splits are the leaf basis portfolio $\mathbf{R}_t^{(J)} = [R_{1,t}^{(J)}, ..., R_{n,t}^{(J)}, ..., R_{J+1,t}^{(J)}]$ and the final latent factor $f_t^{(J)}$. This encodes a set of splits of the characteristics that maximizes the Sharpe ratio of the asset universe.

3.3 Extensions - Boosted Panel Trees

[1]The factor from one P-Tree represents the tangency portfolio of the mean-variance efficient frontier. The mean-variance efficient frontier represents sets of portfolios that have the highest return given a level of risk. The tangency portfolio is the portfolio with the highest Sharpe-ratio among the portfolios that span the mean-variance efficient frontier when investors can invest in a risk-free asset. In this study, the 1-month Swedish T-bill represents the risk-free asset. It is theoretically

[1]The boosted P-Tree model is not used for the main results.

possible to extend the mean-variance efficient portfolio through boosting. The boosting algorithm fits seamlessly into the P-Tree algorithm since it adds an additional step that is similar to the split criterion step in the P-Tree algorithm. If we denote the final factor obtained from the regular P-Tree algorithm with $f_t^{J_1}$, the new optimization function becomes

$$\mathcal{L}(\tilde{c}_{k,m}) = \sqrt{\hat{\mu_F}' \hat{\Sigma}_F^{-1} \hat{\mu}_F} \tag{4}$$

where $F = [f_t^{J_1}, f_t^2]$. After the first boosted tree is complete, the factor set will be $F = [f_t^{J_1}, f_t^{J_2}]$. This procedure can be repeated p times until the final factor set becomes $F = [f_t^{J_1}, \ldots, f_t^{J_p}]$. Given this set of factors, the objective is to find weights of the factors such that

$$\mathcal{L}(F) = \sqrt{\hat{\mu_F}' \hat{\Sigma}_F^{-1} \hat{\mu}_F} \tag{5}$$

is maximized. This represents an all-tree mean-variance efficient portfolio. This approach also uses indirect regularization by optimizing the weights on the factors and not on the leaf portfolios directly. This avoids high-dimensional issues related to optimizing to many weights simultaneously which can lead to poor estimates of the optimal weights and overfitting. Thus, the results are more robust and there is less risk of overfitting.

3.4 Hyperparameter Configuration

The Boosted P-Tree algorithm is governed by several hyperparameters that determine the structure, regularization, and stability of the estimation process. In this study, some hyperparameters followed the same configuration as the model employed by Cong et al. (2025), while other were adjusted to account for the different market structure of the Swedish equity market.

The covariance shrinkage parameter $\gamma = 10^{-4}$ introduces a regularization term to the diagonal of the covariance matrix, ensuring numerical stability when inverting it to compute portfolio weights. A higher value of γ yields smoother, more conservative portfolios by shrinking the covariance matrix toward the identity, whereas a lower value makes the model more sensitive to the estimated covariance structure but may increase instability. The step of the algorithm involving this parameter is

$$\mathbf{w}^{(j)} \propto \left(\hat{\mathbf{E}}[\mathbf{R}_t^{(j)} \mathbf{R}_t^{(j)'}] + \gamma \mathbf{I}\right)^{-1} \hat{\mathbf{E}}[\mathbf{R}_t^{(j)}] \tag{6}$$

At the ensemble level, the covariance factor shrinkage parameter $\lambda = 10^{-5}$ regularizes the covariance matrix of the P-Tree factors themselves. This prevents any single tree from dominating the boosted ensemble and ensures that the combination of factors remains well-conditioned. Smaller values encourage stronger exploitation of factor correlations while larger values promote balance and robustness. Another parameter that controls the boosting is the learning rate η. This parameter controls how much weight each additional factor has in the final factor set. These two hyperparameters work together through

$$\mathbf{w}_F \propto (\hat{\Sigma}_F + \lambda \mathbf{I})^{-1} \hat{\mu}_F$$
$$\mathbf{w}_{final}^{(P)} = [\mathbf{w}_{final}^{(P-1)}, \eta w_F^{(P)}]$$

where $w_F^{(P)}$ denotes the p:th component of $\mathbf{w}_F$

The split thresholds $c_m \in [-0.6, -0.2, 0.2, 0.6]$ define the candidate partition points for firm characteristics. Each potential split is evaluated based on its contribution to the Sharpe ratio, thereby

determining the optimal granularity of cross-sectional partitioning. Narrower thresholds allow finer differentiation among firms but may also increase sensitivity to noise.

To prevent overfitting, a minimum leaf size of three firms per terminal node is imposed. This constraint ensures that each leaf portfolio is estimated from a sufficient cross-section of firms. The parameter value is scaled from the U.S. market to the Swedish context based on the relative number of available monthly stocks. Cross-sectional winsorization is also applied at the 1st and 99th percentiles of firm characteristics and returns, replacing extreme values with these boundary percentiles to limit the influence of outliers and transient price effects.

The number of boosting iterations T is controlled by *num_iter*. It determines how many trees are sequentially grown. Cong et al. (2025) set $T = 9$ boosting iterations. However, due to the limited cross-sectional data of the Swedish market, as can be seen in Figure 1, we empirically determined the optimal number of iterations through out-of-sample validation. Testing with different boosting iterations resulted in more splits, but worse performance and less predictive power. For the Swedish market, the single tree configuration (T = 1, no boosting) provided superior performance. The out-of-sample Sharpe ratio of 0.53, see table 6, could not be beaten by subsequent changes in boosting iterations. In fact, $T \geq 2$ parameters improved in-sample performance, but simultaneously degraded out-of-sample performance (see Appendix 7)

Finally, the maximum tree depth is restricted to 10, limiting the number of consecutive splits per tree. While deeper trees can capture more complex, nonlinear interactions between firm characteristics and returns, excessive depth increases the risk of overfitting. This cap provides a balance between model flexibility and generalization.

Overall, these hyperparameters jointly control the balance between flexibility, stability, and interpretability in the Boosted P-Tree model. The shrinkage parameters (γ, λ and η) govern regularization and numerical stability, while structural parameters such as the split thresholds, leaf size, and tree depth determine the complexity of cross-sectional partitioning.

4 Results

4.1 Data Description

An important aspect of understanding the performance of data-intensive methods such as machine learning approaches is studying the data set. This is an integral part of the research question as one of the key characteristics of the Swedish stock market is the lack of data available to retail investors. Looking at figure 1 we can see that the average number of firms each year is growing. The maximum number of firms for a single time period is roughly 250 and the mean is 153. This shows that the historical coverage is not extensive. To put it into perspective, the total number of firms in the dataset is 575 as can be seen in table 3.

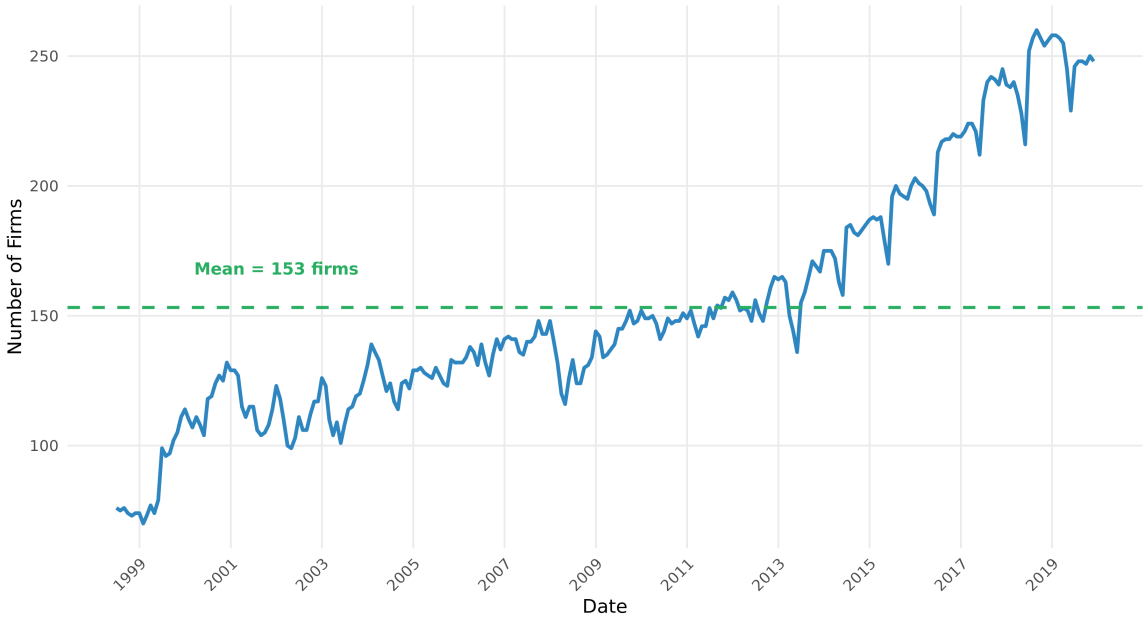


Figure 1: Time series plot of the number of Swedish listed firms included in the dataset over time. The coverage reflects the availability of financial and accounting variables used in the empirical analysis. Variations in coverage occur due to listings, delistings, and missing observations, and therefore determine which firms enter the P-tree model at each point in time.

Table 3: Summary statistics of the dataset used to train the model, including the total number of observations after filtering, the number of unique firms, the temporal span of the data (in months and date range), the number of characteristics, the average cross-sectional coverage (firms per month), and the average temporal coverage (months per firm).

Metric	Value
Observations	39,524
Companies	575
Months	258
Characteristics	33
Date Range	1998-06-30 to 2019-11-30
Avg Companies/Month	153
Avg Months/Company	68.7

The mean coverage per firm is 68.7 months out of a total of 258 months, meaning that most firms are observed for only a fraction of the sample, see figure 2. This intermittent presence may increase noise in the estimated factor returns and exposures, potentially affecting the stability of factor weights.

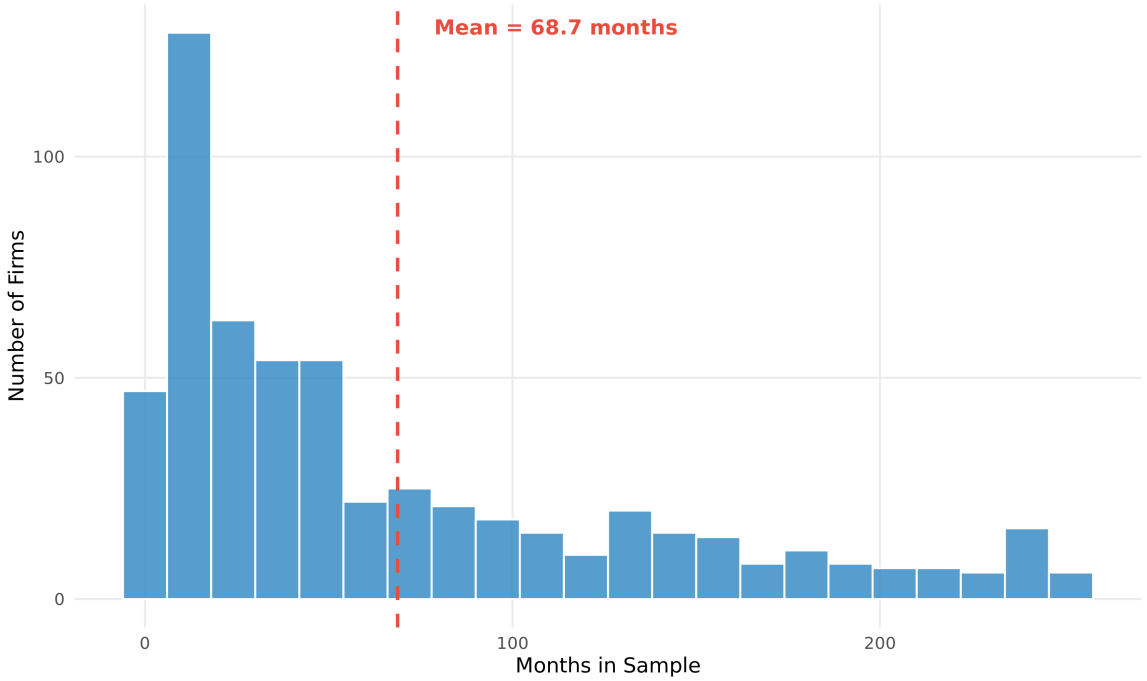


Figure 2: Histogram showing the distribution of firms' temporal coverage, measured as the number of months each firm is present in the dataset. The mean coverage is 68.7 months. The distribution is right-skewed, with the majority of firms exhibiting shorter-than-average coverage.

In comparison, Cong et al. (2025) had a much more comprehensive dataset. They had monthly data from 1981-2020 and the average and median coverage per month was 5265 and 4925 for the first 20 years, and 4110 and 3837 for the latter 20 years. They use 61 characteristics while this study can only use 33. This reflects limited data availability for constructing these variables. Figure 3 shows the difference in total number of observations and number of characteristics.

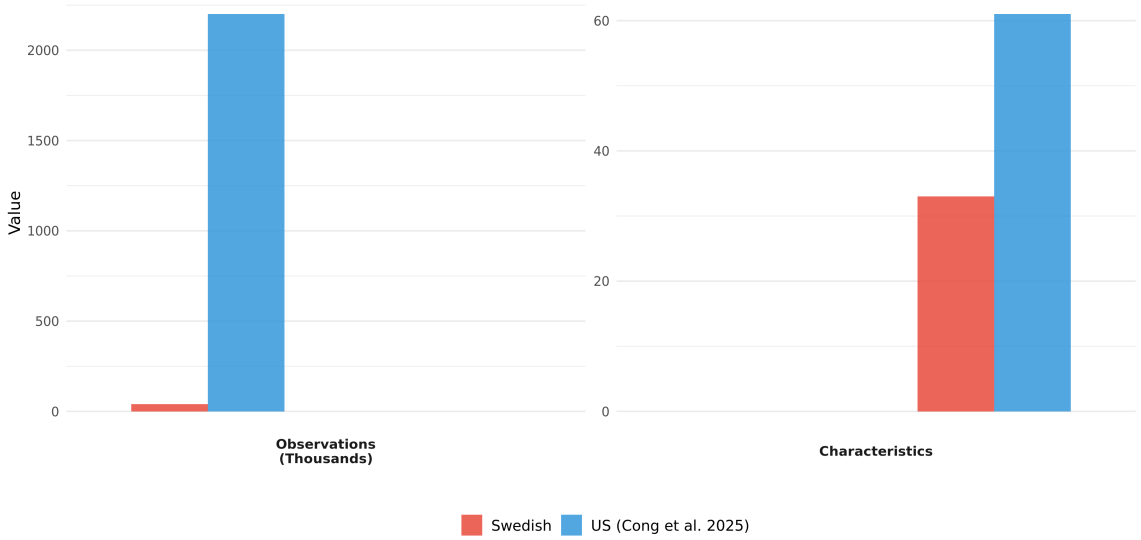


Figure 3: Bar-plot comparing the data availability with the original paper. The plots shows comparisons of the total number of observations and the number of characteristics used. The original paper had approximately 2,250,000 observations and 61 characteristics, while this study had 39,524 observations and 33 characteristics.

The correlation of the explanatory variables is always important to consider for any prediction model, see Appendix 7. The univariate R^2 is shown in table 5. Some variables are highly correlated with a number of different variables, while show little correlation. This is due to many variables measuring similar characteristics or are transformation of other variables. For example, operating accruals (ACC) and percent accruals (PCTACC) are highly correlated ($\rho = 0.89$) as they are linked through one transformation ($PCTACC = \frac{ACC}{Assets}$). However they have little correlation with the other variables. On the other hand, return on equity (ROE) is highly correlated with a number of other variables that also measures other forms of profitability ratios. Variables with a correlation factor $|\rho| > 0.7$ are shown in table 4.

Table 4: Highly correlated characteristic pairs with $|\rho| > 0.7$. Many variables are derived from common underlying measures, which naturally induces correlation. 16 out of the 33 characteristics have a correlation coefficient above 0.7.

Characteristic 1	Characteristic 2	ρ
CHCSHO	NI	1.000
ROA	ROE	0.954
ATO	GMA	0.906
GMA	SP	0.900
CASH	CASHDEBT	0.899
EP	ROA	0.887
OP	RNA	0.880
ACC	PCTACC	0.872
EP	ROE	0.870
ATO	SP	0.831
CFP	EP	0.805
CFP	ROA	0.745
CFP	ROE	0.708
EP	PM	0.707
PM	ROA	0.701

Note: CHCSHO and NI both measure changes in shares outstanding but use different transformations: CHCSHO = (shares - shares_lag12) / shares_lag12 (percentage change), while NI = log(shares / shares_lag12) (log change). Their perfect correlation reflects the fact that they capture the same underlying information about equity issuance activity.

There are a number of profitability related variables that are highly correlated. Highly correlated variables can generally cause estimates to become unstable. Many of these act as a proxy for the performance of the firm. As such, when a firm is performing well most of the performance related variables will have high values and the opposite when the performance is low.

The univariate R^2 value is a measure of how much the variance in the return can be explained by a single variable. Looking at table 5 the top three variables with the highest univariate R^2 are all momentum related. This suggest that winners tend to stay winners. This is also supported by the annual spread which is the return difference between portfolios formed on the extremes of the characteristic. However, while the individual univariate R^2 values are small with the largest being smaller than 0.0032, the spreads are still large for some of the variables, the characteristic can still generate economically meaningful trading profits.

Table 5: Univariate predictive power of characteristics, measured by R^2, along with the annual return spread between the top 20 and bottom 5 stocks. Although individual variables have limited explanatory power, differences between extreme quintiles can produce substantial annual spreads. The t-stat is for the annual spread. 13 characteristics have a statistically significant annual spread.

Characteristic	N	R^2	Annual Spread (%)	t-stat	Sig
MOM12M	32,609	0.003181	27.37	10.20	***
MOM6M	35,377	0.002450	24.17	9.32	***
MOM60M	17,671	0.000293	7.53	2.27	**
ROA	39,243	0.000214	7.29	2.90	***
PM	31,640	0.000210	7.17	2.58	***
EP	39,280	0.000206	7.19	2.84	***
MOM36M	23,395	0.000177	6.20	2.03	**
RNA	39,087	0.000154	6.22	2.46	**
CFP	26,634	0.000144	6.36	1.96	**
NI	32,552	0.000129	-5.51	-2.05	**
CHCSHO	32,551	0.000129	-5.51	-2.05	**
ROE	39,139	0.000127	5.63	2.23	**
OP	39,275	0.000113	5.33	2.11	**
CHTX	26,201	0.000113	4.88	1.72	*
PCTACC	22,302	0.000092	4.96	1.43	
SEAS1A	31,686	0.000074	4.17	1.54	
LEV	39,387	0.000043	-3.28	-1.30	
ACC	22,881	0.000035	3.03	0.89	
HIRE	30,700	0.000029	-2.70	-0.95	
GRLTNOA	25,234	0.000025	2.32	0.79	
GMA	39,377	0.000001	-0.45	-0.18	
CASH	39,391	0.000000	-0.33	-0.13	
SP	39,415	0.000000	0.11	0.04	
NOA	28,988	0.000000	-0.00	-0.00	
SGR	28,365	0.000000	0.00	0.00	

Note: *** p<0.01, ** p<0.05, * p<0.10. Only non-zero observations used.
Annual Spread: Annualized return difference between highest and lowest ranked stocks (long-short portfolio).
Median R^2 across all 33 characteristics: 0.000043

4.2 Main Model Performance

To evaluate the performance of the P-Tree algorithm in the Swedish equity market, we implemented three distinct testing scenarios, based on the methodology from Cong et al. (2025). This multi-scenario approach allows us to assess both in-sample fit and out-of-sample predictive power, while testing for robustness across different time periods. Table 6 summarizes the risk-adjusted performance of the Single P-Tree long-short portfolios across all three scenarios. The results demonstrate that the algorithm is capable of generating significant excess return in the Swedish market, though performance varies considerably by time period. The weights and splits for the scenarios are shown in table 7.

Table 6: Performance of the P-Tree model, including annualized return, volatility, Sharpe ratio, β, alphas, and out-of-sample R^2. The sample is split temporally into two halves, with the value-weighted market portfolio as the benchmark.

Scenario	Ann. Ret (%)	Vol (%)	Sharpe	β	CAPM α (%, t)	FF3 α (%, t)	R^2 (%)
Market (VW)	10.86	19.01	0.64	1.00	–	–	–
Scenario A	9.80	10.78	0.92	0.03	9.68*** (4.45)	9.61*** (4.43)	20.20
Scenario B (Train)	9.69	11.04	0.90	-0.08	10.21*** (3.36)	10.11*** (3.43)	–
Scenario B (Test)	2.76	5.37	0.53	-0.02	3.08** (2.07)	3.21** (2.02)	14.61
Scenario C (Train)	10.46	7.18	1.43	0.05	9.59*** (5.58)	9.63*** (5.35)	–
Scenario C (Test)	3.48	7.82	0.48	0.01	3.71 (1.33)	3.39 (1.21)	25.05

Notes: Ann. Ret and Vol are annualized percentages. Alphas are annualized (%) with Newey–West t-statistics (12 lags) shown in parentheses. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. R^2 is out-of-sample.

Scenario A is a full sample analysis. The model trains and evaluates the P-Tree algorithm on the complete sample period from June 1998 to December 2019. This scenario serves as a benchmark for the model's performance when all available data is utilized. The full-sample analysis demonstrates the maximum capability of the P-Tree in the Swedish market and identifies which characteristics are most influential over the entire period. Since the model is both trained and evaluated on the same dataset, Scenario A does not provide insight into out-of-sample predictive power. In this full-sample analysis, the Single P-Tree model achieves an annualized Sharpe ratio of 0.92. The portfolio generates a statistically significant CAPM alpha of 9.68% (t = 4.45) and a Fama–French three-factor alpha of 9.61% (t = 4.43), as shown in Table 6. The model splits stocks based on asset turnover (ATO), with the threshold at the 40th percentile, allocating approximately 75% of the weight to the top 60th percentiles and shorting the bottom 40th percentiles with 25% of the weight (see Table 7). These results indicate that the P-Tree algorithm can successfully fit historical data and extract a profitable signal from the selected characteristics. The obtained Sharpe ratio is greater than the market's of 0.64. However, because the full dataset is used for both training and evaluation, these metrics should be interpreted as an upper bound on the model's fitting capability rather than a measure of true predictive skill.

Scenario B employs a past-predicting-future approach. The model implements a traditional out-of-sample test by dividing the sample chronologically at the end of 2009. The model is trained on data from June 1998 to December 2009 and evaluated on the subsequent period from January 2010 to December 2019. This forward-validation approach tests whether patterns learned from historical data can predict future returns, which is the most practically relevant question for investors. By construction, the model has no access to future information during training, making this the primary test of real-world applicability. In Scenario B, the model splits on asset turnover at the 40th percentile, shorting the bottom 40th percentiles with a weight of approximately 33% and allocating roughly 67% to a long position in the top 60th percentiles. While the in-sample performance is strong (Sharpe ratio = 0.90, CAPM α = 10.21% (3.36), FF3 α = 10.11% (3.43)), the out-of-sample results are weaker (Sharpe ratio = 0.53, CAPM α = 3.08% (2.07), FF3 α = 3.21% (5.35)), falling below the market Sharpe ratio. This decline in out-of-sample performance indicates that the model captures some noise in the training data and exhibits overfitting.

Scenario C employs a future-predicting-past approach. This provides a robustness check by reversing the train-test split: the model is trained on data from January 2010 to December 2019 and tested on the earlier period from June 1998 to December 2009. This approach serves two purposes.

First, it tests for look-ahead bias by evaluating whether the performance observed in Scenario B could be due to fortuitous timing or the selection of a favorable test period. Second, it assesses whether the relationships between characteristics and returns are persistent across different market regimes. If both Scenarios B and C produce similar out-of-sample results, this would provide strong evidence that the P-Tree captures fundamental, time-invariant patterns rather than period-specific anomalies. In Scenario C, the model splits based on the cash-to-debt ratio, shorting the bottom 20th percentile with a weight of approximately 37% and allocating roughly 63% to the top 80th percentile. The in-sample performance is the strongest among all scenarios (Sharpe ratio = 1.43, CAPM $\alpha = 9.59\%$ (5.58), FF3 $\alpha = 9.63\%$ (5.35)), but the out-of-sample results are weaker (Sharpe ratio = 0.48, CAPM $\alpha = 3.71\%$ (1.33), FF3 $\alpha = 3.39\%$ (1.21)), with the alpha no longer statistically significant. This decline in out-of-sample performance indicates that the model captures some noise in the training data and exhibits overfitting.

Table 7: Portfolio weights and splitting criteria from the P-Tree model for the in-sample and out-of-sample tests. The model performs a single split in all scenarios. For the in-sample and first-half training, the split is based on asset turnover, while for the second-half training, the split is based on the cash-to-debt ratio.

Scenario	Split Variable	Portfolio	Threshold	Weight (%)
A: Full Sample	ATO	Low ATO ($<$ -0.20)	-0.20	-25.30%
A: Full Sample		High ATO ($\geq$ -0.20)	-0.20	74.70%
B: Train (1998-2009)	ATO	Low ATO ($<$ -0.20)	-0.20	-33.06%
B: Train (1998-2009)		High ATO ($\geq$ -0.20)	-0.20	66.94%
C: Train (2010-2019)	CASHDEBT	Low CASHDEBT ($<$ -0.60)	-0.60	-36.55%
C: Train (2010-2019)		High CASHDEBT ($\geq$ -0.60)	-0.60	63.45%

Note: Weights represent the tangency portfolio allocation across leaf portfolios, optimized to maximize the Sharpe ratio. Within each leaf, stocks are value-weighted by market capitalization.

The cumulative return profiles in Figure 4 illustrate how each portfolio performed over the sample period. Although the in-sample P-Tree portfolio achieves a higher Sharpe ratio than the market, the market portfolio still delivers a higher cumulative return. This occurs because the in-sample portfolio exhibits substantially lower volatility, boosting its risk-adjusted performance despite generating lower overall returns. This comparison highlights the fundamental trade-off between risk and return. A well-diversified portfolio can achieve smoother performance and a higher Sharpe ratio, but higher long-run returns generally require bearing greater risk.

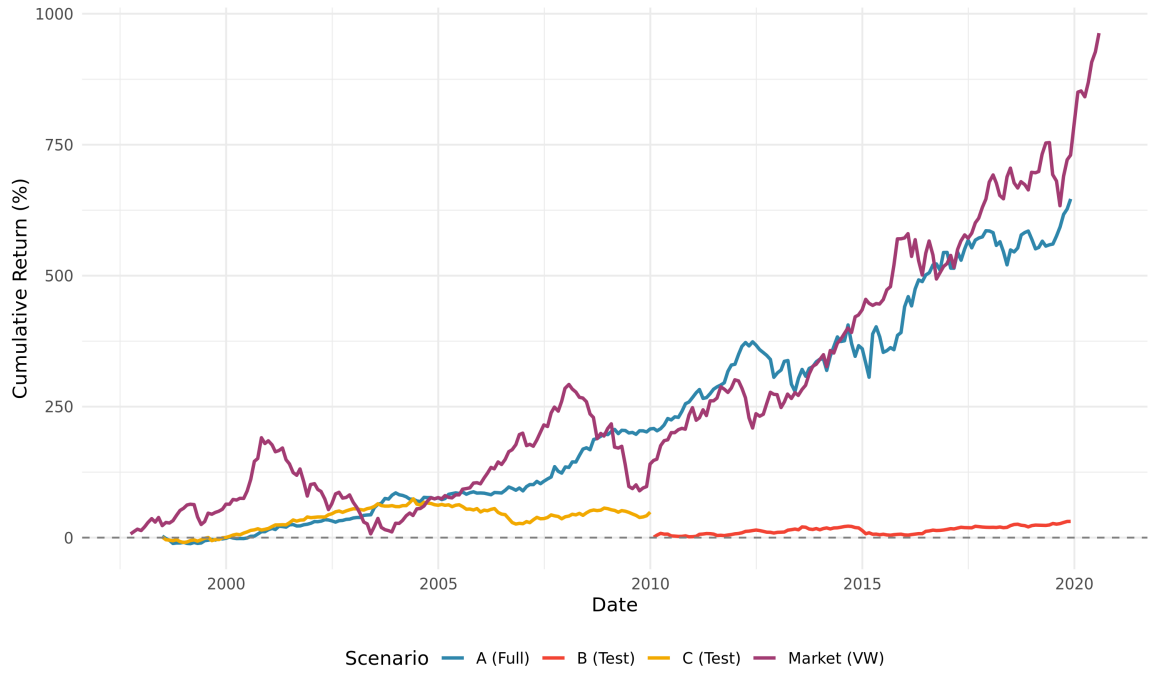


Figure 4: Cumulative returns of the value-weighted market portfolio, the in-sample portfolio, and the out-of-sample test portfolios. Cumulative return is calculated as $\prod_{i=1}^{T}(1+r_i)$. Even though the in-sample test achieved a higher Sharpe-ratio than the market portfolio, the cumulative return of the market portfolio is greater.

4.3 Macroeconomic Regimes

Figure 5 shows the risk-free rate, market return and market risk premium as time series. The rest of the macroeconomic factors are in the Appendix 7. The market, size, and value factors are obtained from the Swedish House of Finance's Fama–French dataset. Inflation data are sourced from Statistics Sweden Statistics Sweden (SCB) (2025), which reports annual inflation based on the Consumer Price Index (CPI) with 1980 as the base year. Annualized market volatility is computed from monthly market index returns. The risk-free rate, measured using the Swedish 1-month T-bill, has been near zero since March 2015. This follows the Swedish central bank's decision to cut the policy rate into negative territory, which caused short-term interest rates to fall and, consequently, reduced the yields on short-term government securities. The market index is the SIX return index which is a value weighted index of all stocks listed at the Stockholm Stock Exchange and includes reinvested dividends. This serves as a natural benchmark as the P-Tree algorithm constructs value-weighted leaf basis portfolios.

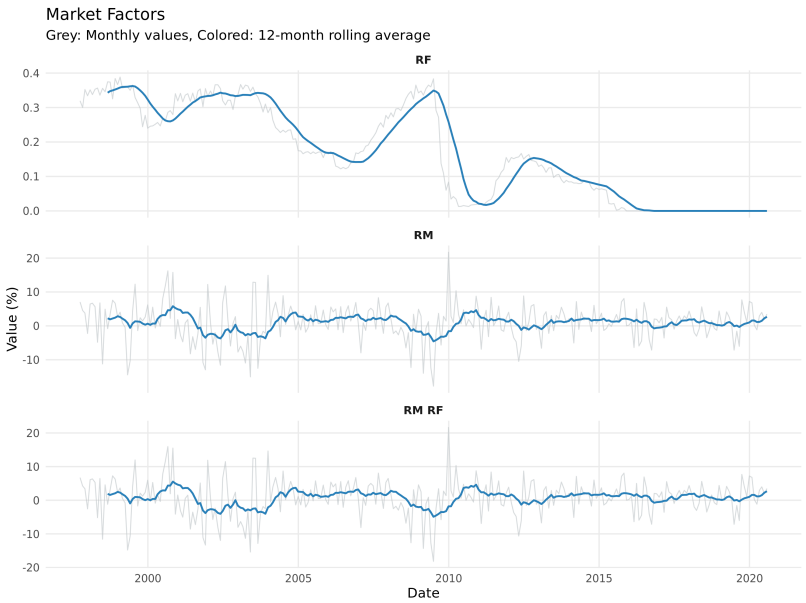


Figure 5: Time series of the risk-free rate, market returns, and the market risk premium. Monthly data points are shown in gray, with a 12-month rolling average in blue. All values are expressed as percentages on the y-axis.

5 Empirical Analysis and Discussion

5.1 Model Performance and Robustness

To understand the drivers of model performance across different scenarios, the structure of the P-Trees shown in Table 7 is analyzed. In all scenarios, the model is configured to allow a single split, resulting in two terminal leaves. For the in-sample and out-of-sample Scenario B results, the model selects asset turnover at the 40th percentile as the optimal split, while in Scenario C, the split occurs on the cash-to-debt ratio at the 20th percentile. This setup highlights that, for the given universe and sample period, the dominant predictive signal is concentrated in one characteristic rather than in higher-order interactions or complex nonlinear structure. This can be observed with the decrease in the out-of-sample performance when the model complexity increases, (see Appendix 7). One notable finding is that four out of five scenarios produced statistically significant α at 10% confidence level under both the CAPM and the Fama–French three-factor model, with the exception being the scenario C's out-of-sample result. This suggests that, although the model does not consistently outperform the market portfolio in terms of traditional risk-adjusted metrics such as the Sharpe ratio in every scenario, it is able to capture return premiums that are not explained by the included risk factors, namely the market, size (SMB), and value (HML) factors. The near-zero beta values help explain these results. Across all scenarios, the estimated Beta values are effectively zero (Table 6), which indicates that the strategy's returns are uncorrelated with the broader market movement (Market Neutrality). A likely interpretation is that the alpha generated by the model is entirely derived by its ability to identify idiosyncratic mispricing. The P-Tree could thus play an integral role in an investor's market portfolio. By reducing overall portfolio volatility, independent of market rallies or crashes, the P-Tree strategy can provide diversification benefits.

Figure 4 shows that the in-sample cumulative return and table 6 performance metrics of the different scenarios. However, although Scenario A exhibits the highest Sharpe ratio, the market portfolio generates the largest cumulative return. This highlights the fundamental trade-off between risk and return. The market achieves higher total returns, but only by taking on substantially more volatility. In contrast, the P-Tree portfolio prioritizes stability and delivers superior risk-adjusted performance. The model is designed not to maximize absolute returns, but to produce returns in the safest and most consistent manner.

The best results obtained when using a single split indicates that, given the limited size and coverage of the dataset, additional partitioning does not yield meaningful improvements in predictive accuracy. This interpretation is reinforced by the results obtained when the model is allowed to perform up to two and ten splits (see Appendix 7). Although the in-sample performance improves reflected in higher Sharpe ratios and larger estimated alphas, the corresponding out-of-sample results deteriorate substantially, suggesting that deeper trees begin to fit noise rather than genuine return-predictive structure.

Importantly, the P-Tree framework enables the user to control the level of model complexity, reducing the risk of overfitting when the dataset is sparse. Under sparse data conditions, strong regularization and a conservative choice of the number of splits encourage a parsimonious [2] specification, while richer datasets may justify allowing deeper trees or using boosting to capture more complex nonlinear interactions. In the present application, however, additional splits reduce both the Sharpe ratio and the estimated alphas, indicating that the model does not uncover new, independent sources of predictive power. This implies that the dataset does not contain weak but orthogonal predictors that could support greater model complexity; instead, additional characteristics appear either redundant or noisy. Consequently, the underlying return-predictive structure is relatively simple, and increasing model depth offers no performance advantage. By giving the user control over complexity, the model is well-suited for both data-sparse and data-rich environments, highlighting its practical potential for portfolio construction and systematic investing.

5.2 Characteristic Selection and Macroeconomic Context

The selection of splitting characteristics must be interpreted within the macroeconomic context of each sample period. As illustrated in Figure 7, the risk-free rate in Sweden declined significantly starting around 2015, eventually reaching zero. This monetary policy was intended to stimulate economic activity and incentivize investment. Interestingly, the model's selection of Cash-to-Debt as the primary splitting criterion in Scenario C (out-of-sample) suggests that, despite an environment favorable to leverage, the market priced a premium on financial flexibility and balance sheet strength.

In contrast, during the earlier period (1998-2009), the model identified Asset Turnover as the single determining factor. This period was marked by two major global economic downturns (Dot-com bubble and Global Financial Crisis) which significantly impacted export-oriented Swedish equities. The selection of Asset Turnover implies that during these volatile regimes, the market favored firms capable of efficiently generating revenue from their asset base. Furthermore, the model's consistent preference for fundamental accounting ratios (Asset Turnover, Cash-to-Debt) over price-based signals (such as Momentum) indicates that fundamental metrics were more robust drivers of returns in the Swedish market during these regimes.

[2] Parsimonious: simple, economical, not overly complex

5.3 Data Availability and Comparison to Original Paper

While data for Swedish equities is available, its coverage and depth differ substantially from the US data used in the original P-Tree study by Cong et al. (2025). To approximate half of the characteristics used in the original study, data aggregation from multiple sources was required, and several key characteristics remained inaccessible. Notably, the original P-Tree paper relies heavily on analyst revision and the characteristic standardized unexpected earnings (SUE) which frequently appear as the primary splits in their models. Comprehensive historical analyst forecast data for the Swedish market is not readily available for the full sample period. Consequently, our model relies on "second-best" proxies derived from lagged accounting data, which likely diminishes predictive power.

Furthermore, there is a significant disparity in market size between the US market, comprising thousands of liquid stocks, and the Swedish market, which averages approximately 153 firms per month in our filtered sample. While this sample size is sufficient for traditional analysis, it proves limiting for the P-Tree algorithm. A multi-layer tree (e.g., a 3-layer depth) would result in leaf nodes containing fewer than 10 stocks, thereby destroying portfolio diversification and increasing idiosyncratic risk.

Compared to Cong et al. (2025), whose model achieved out-of-sample Sharpe ratios in the range of 3.12–4.35 and α values statistically significant at the 1% level for all factor combinations tested (see Appendix 7), the results in this study are less pronounced in terms of Sharpe ratios and α significance. Despite this, the results demonstrate that meaningful predictive performance can still be obtained without overfitting the data. This highlights the robustness of the P-Tree methodology even in a smaller, less liquid market.

5.4 Parameter Stability

Over the course of the analysis, an extensive sensitivity analysis was conducted involving hyperparameter tuning, alternative lagging structures, and varying data filtration criteria. A critical finding from this process is that the P-Trees exhibited high sensitivity to minor changes in the input data and configuration. Experiments involving automatic hyperparameter tuning led to unstable outcomes. Parameters would exactly fit the noise in the training data, instead of finding a global stable optimum, and fail to generalize out-of-sample. A behavior which can be attributed in large parts to the small market size. Similarly, modifications to the lagging methodology included altering the publication lag for accounting variables. These resulted in significant shifts in the final tree structure and performance metrics (still with one split). Notably, the results are heavily influenced by the data filtration step. For instance, adjusting the threshold for permissible missing values per firm from 4 to 5, changes the split characteristics. In a large market like the US, such a change would be negligible. However, in the Swedish market, where the cross-section averages only 153 firms, the exclusion or inclusion of a small number of firms has a much more significant impact.

5.5 Practical Implementation

Another important factor that should be considered when applying this methodology in a real world scenario are market frictions. High turnover strategies often see their profits eroded entirely by the bid-ask spread. Especially in smaller, less liquid markets such as the Swedish equity market, transaction costs will be higher. To achieve more realistic returns, implementing transaction costs for frequent rebalancing would likely have diminished, if not completely offset, the alphas. Another aspect that differs in practice versus our implementation is when a stock delists from the exchange.

The model avoids survivorship bias by dynamically dropping stocks from the investable universe when a firm delists. However, in practice these exits can be costly due to the increase in the bid-ask spread as there will be a large number of investors selling the stock simultaneously.

6 Conclusion

The integration of machine learning into portfolio optimization has gained substantial attention in modern finance. Machine learning methods excel at processing large-scale panel data and identifying complex, non-linear relationships that traditional models, such as linear regressions, are unable to capture. While other machine learning models, including deep neural networks, can also uncover intricate interactions among firm characteristics, the novel Panel Tree (P-Tree) algorithm offers several important advantages.

First, unlike many black-box models, the P-Tree algorithm optimizes a user-specified objective function. In the context of portfolio optimization, this objective can be set to the Sharpe ratio, enabling the model to directly search for return–risk–optimal splits. Second, the factors produced by the P-Tree have clear economic interpretation as the selected splits and associated weights highlight fundamental drivers of returns. This not only provides economic insight into the cross-section of expected returns but also allows researchers to diagnose potential model misspecification when the learned factors deviate from established empirical patterns.

In contrast, many other machine learning models lack transparency, offer no graphical representation of decision logic, and are not grounded in economic structure. The P-Tree algorithm therefore combines the predictive flexibility of machine learning with interpretability and economic meaning, making it a compelling tool for empirical asset pricing.

While machine learning methods excel when trained on rich and extensive panel datasets, their flexibility also makes them susceptible to overfitting in environments with limited data. In smaller markets with constrained historical depth and cross-sectional coverage, models may capture noise rather than genuine economic signals, which can reduce out-of-sample predictive performance. This pattern is visible in the results with the out-of-sample Sharpe ratios falling below that of the market portfolio, reflecting the challenges of data sparsity.

Despite this, the P-Tree demonstrates notable strengths in these environments. The portfolio is nearly market-neutral, with $\beta \approx 0$, yet it is still is able to produce statistically significant positive α across most specifications under the Fama–French three-factor model. While the strategy may not surpass a simple buy-and-hold market portfolio over the long term, it delivers investable, low-beta return streams that provide diversification and reduce exposure to broad market movements. These features are particularly valuable in the short run, during periods of heightened volatility, offering investors a reliable tool to extract returns from the most informative characteristics without over-relying on data-intensive, complex models. The adaptability of the P-Tree is a decisive advantage. This flexibility allows the model to function effectively in smaller and less liquid markets where data is constrained, such as the Swedish stock market. In contrast, many other machine learning methods struggle in these settings because they become overly complex or prone to noise-fitting. As datasets grow richer and more comprehensive in the future, the P-Tree model is well-positioned to scale its structure, bridging the gap between sparse-data environments and sophisticated machine learning methods, and offering a powerful tool for portfolio optimization.

References

Ammann, M., & Steiner, P. (2023). Factor premiums in the nordic stock markets. *Journal of International Financial Markets, Institutions and Money*, *68*, 101255.

Bali, T. G., Engle, R. F., & Tang, Y. (2024). Is there a market-neutrality premium? *Journal of Financial Economics*, *152*, 103846.

Bates, T. W., Kahle, K. M., & Stulz, R. M. (2023). Why do u.s. firms hold so much more cash than they used to? *Journal of Finance*, *78*(5), 2595–2645.

Bryzgalova, S., Pelger, M., & Zhu, J. (2023). Forest through the trees: Building cross-sections of stock returns with gradient boosting. *Journal of Finance*, *78*(3), 1411–1458.

Cong, L. W., Feng, G., He, J., & He, X. (2025). Growing the efficient frontier on panel trees. *Journal of Financial Economics*, *167*, 103873. https://doi.org/10.1016/j.jfineco.2024.103873

Didisheim, A., Koijen, R. S. J., & Yogo, M. (2024). *High-throughput asset pricing* (Working paper). University of Chicago & Princeton University.

Drobetz, W., & Otto, T. (2021). Empirical asset pricing via machine learning: Evidence from the european stock market. *Journal of Asset Management*, *22*(7), 507–538.

Freyberger, J., Neuhierl, A., & Weber, M. (2024). Dissecting anomaly discovery. *Journal of Financial Economics*, *151*, 103745.

Ghandar, A., Michalewicz, Z., & Zurbruegg, R. (2016). The relationship between model complexity and forecasting performance for computer intelligence optimization in finance. *International Journal of Forecasting*, *32*(3), 598–613.

Griffin, J. M. (2002). Are the fama and french factors global or country specific? *Review of Financial Studies*, *15*(3), 783–803.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Hou, K., Xue, C., & Zhang, L. (2022). Replicating anomalies. *Review of Financial Studies*, *35*(2), 707–746.

Lseg. (2025). Retrieved November 28, 2025, from https://www.lseg.com/en/data-analytics

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

Martin, I. W. R., & Nagel, S. (2021). Market efficiency in the age of big data. *Journal of Financial Economics*, *141*(1), 3–32.

Simutin, M. (2010). Excess cash and stock returns. *Journal of Financial and Quantitative Analysis*, *45*(4), 765–790.

Soliman, M. T. (2008). The use of dupont analysis by market participants. *The Accounting Review*, *83*(3), 823–853.

Statistics Sweden (SCB). (2025). Cpi, annual changes (inflation rate) [Accessed: 2025-12-04].

Swedish House of Finance. (2025a). *Fama–french factors for the swedish market* [Dataset provided by the Swedish House of Finance]. Retrieved November 30, 2025, from https://www.hhs.se/en/houseoffinance/data-center/fama-french-factors/

Swedish House of Finance. (2025b). *Finbas* [Dataset provided by the Swedish House of Finance]. Retrieved November 28, 2025, from https://www.hhs.se/en/houseoffinance/data-center/finbas-stocks/

Swedish House of Finance. (2025c). *Serrano* [Dataset provided by the Swedish House of Finance]. Retrieved November 28, 2025, from https://www.hhs.se/en/houseoffinance/data-center/serrano/

7 Appendix

7.1 AI Usage

The following AI tools were used in this project:
GitHub Copilot / Claude: Code assistance, debugging code, refactoring suggestions, and documentation formatting
Usage in the writing: Grammar checking and improvements, code refactoring suggestions, and LaTeX table generation.
All analytical decisions, methodology choices, and interpretations are the authors' own work.

7.2 Supplementary Tables and Figures

Table 8: The full variable names, abbreviations and total coverage in the dataset. These are all variables included in the P-Tree algorithm.

Variable	Abbrev	Coverage (%)
1-Month Momentum	MOM1M	99.7
12–2 Month Momentum	MOM12M	82.5
36–13 Month Momentum	MOM36M	59.2
60–13 Month Momentum	MOM60M	44.7
6–2 Month Momentum	MOM6M	89.5
Asset Growth	AGR	85.4
Asset Turnover	ATO	85.2
Book-to-Market	BM	99.4
Cash to Assets	CASH	99.7
Cash to Debt	CASHDEBT	98.6
Cash-Flow-to-Price	CFP	67.4
Change in Profit Margin	CHPM	66.2
Change in Shares Outstanding	CHCSHO	82.4
Change in Tax Expense	CHTX	66.3
Earnings-to-Price	EP	99.4
Employee Growth	HIRE	77.7
Gross Profitability	GMA	99.6
Growth in LT NOA	GRLTNOA	63.8
Leverage	LEV	99.7
Long-term Debt Growth	LGR	60.9
Market Equity	ME	99.7
Net Equity Issuance	NI	82.4
Net Operating Assets	NOA	73.3
Operating Accruals	ACC	57.9
Operating Profitability	OP	99.4
Percent Accruals	PCTACC	56.4
Profit Margin	PM	80.1
Return on Assets	ROA	99.3
Return on Equity	ROE	99.0
Return on NOA	RNA	98.9
Sales Growth	SGR	71.8
Sales-to-Price	SP	99.7
Seasonality (1Y ago)	SEAS1A	80.2

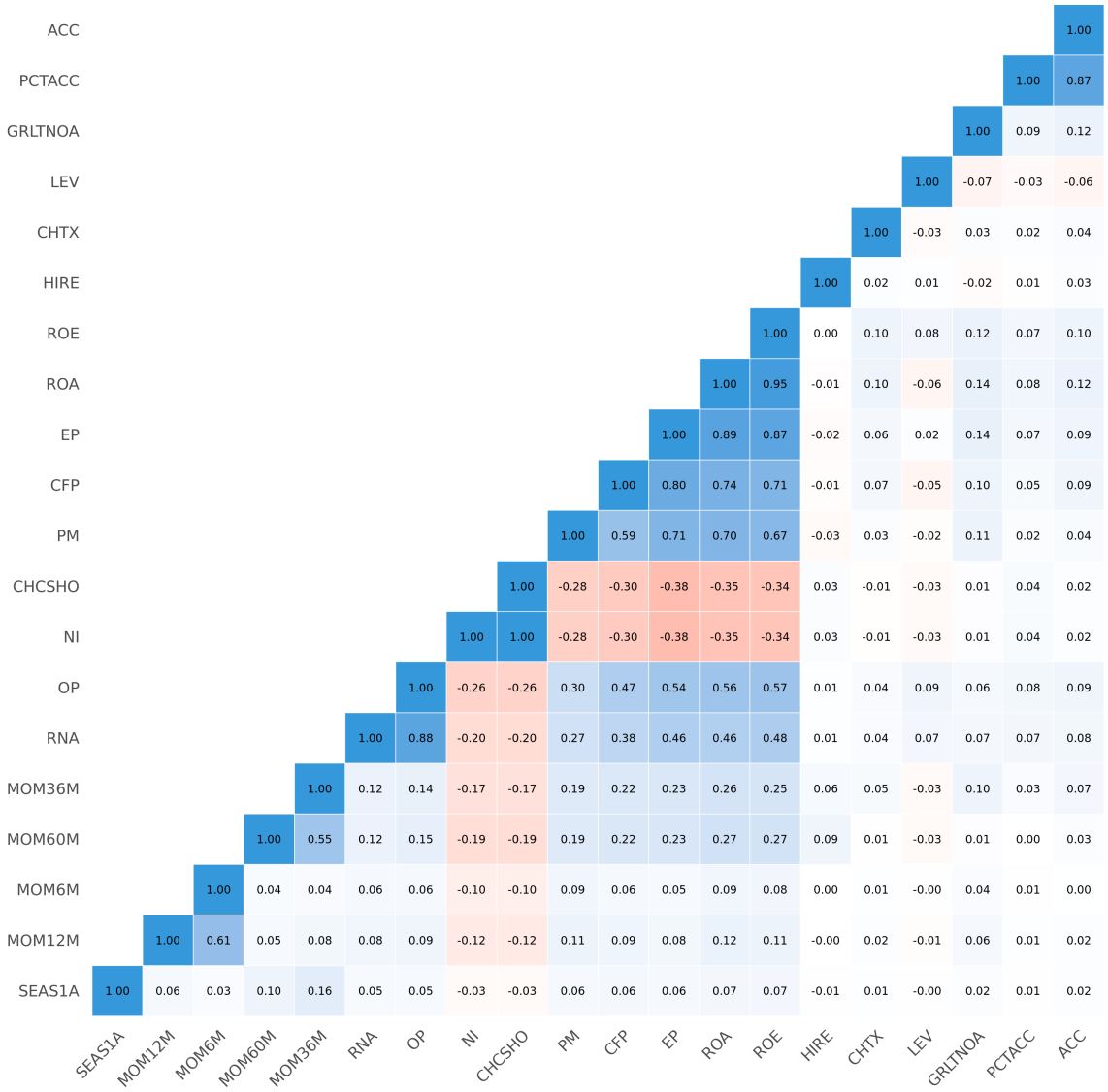
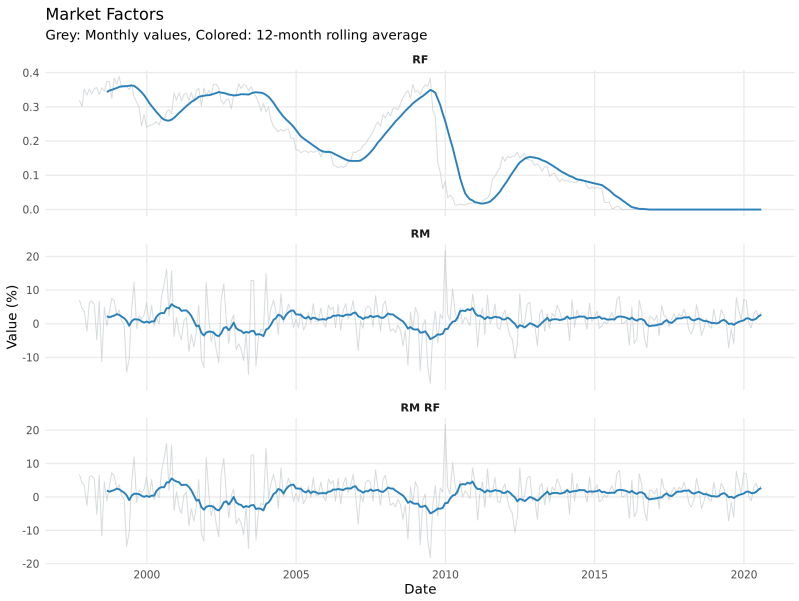
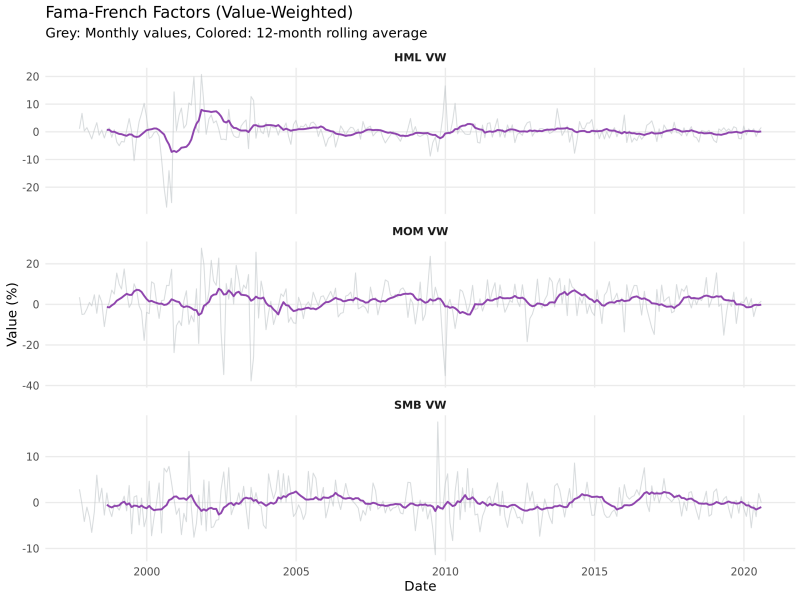


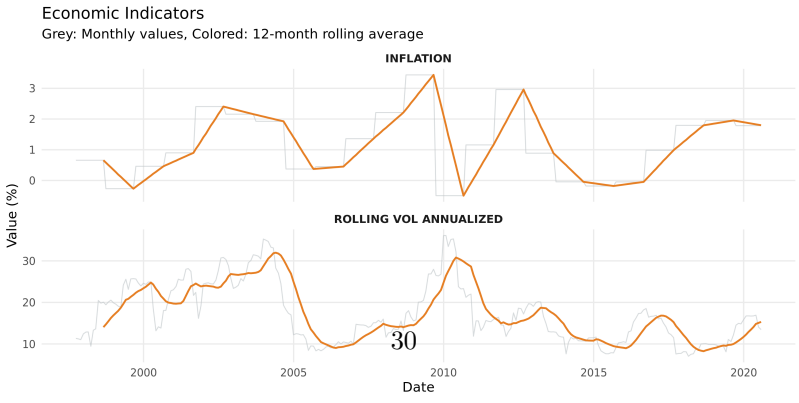
Figure 6: Correlation table for the top characteristics with the highest univariate R^2 value. Some variables are highly correlated with a number of other variables.



(a) Market Overview



(b) Factor Returns



(c) Macro Indicators

Figure 7: Full time series for all the macroeconomic variables used, expressed in percentages. The gray lines represent the raw data points while the colored lines are rolling averages over 12 months.

Table 9: P-Tree model performance when the maximum number of splits is set to ten. Scenario A and B produce nine splits and scenario C produces ten splits. The in-sample results are better compared to when only one split is allowed, but the out-of-sample results are worse indicated by the lower Sharpe ratio and less significant α. This shows that deep trees tend to overfit as they capture noise rather than true relationships.

Scenario	Ret (%)	Vol (%)	Sharpe	β	CAPM α (%, t)	FF3 α (%, t)	R^2 (%)
Market (VW)	10.9	19.0	0.64	1.00	–	–	–
A	10.4	5.5	1.88	0.02	9.73*** (6.84)	9.73*** (7.02)	–
B (Train)	12.1	6.4	1.90	-0.01	11.09*** (4.82)	10.65*** (4.90)	–
B (Test)	0.9	4.7	0.19	-0.02	0.79 (0.79)	0.86 (0.85)	0.5
C (Train)	11.1	2.9	3.88	0.04	10.61*** (19.36)	10.63*** (18.40)	–
C (Test)	2.8	6.5	0.43	-0.03	3.80 (1.61)	3.47 (1.37)	1.4

Notes: Ret and Vol are annualized. Alphas are annualized (%) with Newey–West t-stats (12 lags). Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. R^2 is out-of-sample.

Table 10: P-Tree model performance when the maximum number of splits is set to two. All scenarios produce two splits. The in-sample results are better compared to when only one split is allowed, but the out-of-sample results are worse indicated by the lower Sharpe ratio and less significant α. This shows that deep trees tend to overfit as they capture noise rather than true relationships.

Scenario	Ret (%)	Vol (%)	Sharpe	β	CAPM α (%, t)	FF3 α (%, t)	R^2 (%)
Market (VW)	10.9	19.0	0.64	1.00	–	–	–
A	10.9	9.7	1.12	0.07	8.19*** (4.21)	7.84*** (4.26)	–
B (Train)	13.2	11.4	1.15	-0.05	13.55*** (5.00)	13.55*** (4.95)	–
B (Test)	1.3	5.8	0.22	0.02	0.35 (0.27)	0.45 (0.37)	0.3
C (Train)	9.4	5.2	1.83	0.04	9.10*** (5.10)	9.17*** (5.38)	–
C (Test)	3.4	6.8	0.50	-0.02	3.11 (1.00)	2.71 (0.85)	0.5

Notes: Ret and Vol are annualized. Alphas are annualized (%) with Newey–West t-stats (12 lags). Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. R^2 is out-of-sample.

Table 11: Scenario B: OOS P-Tree Performance (Cong et al. 2025)

Model	Sharpe Ratio	α_{CAPM}	α_{FF5}	α_{Q5}	α_{RP5}	α_{IP5}
P-Tree1	3.23	1.35***	1.31***	1.23***	1.04***	0.93***
P-Tree1-5	3.41	1.02***	1.00***	0.95***	0.77***	0.62***
P-Tree1-10	3.21	0.95***	0.94***	0.89***	0.74***	0.56***
P-Tree1-15	3.12	0.89***	0.89***	0.83***	0.69***	0.48***
P-Tree1-20	3.13	0.85***	0.84***	0.78***	0.66***	0.49***

Source: Cong et al. (2025), Table 7, Panel B2. Results are for out-of-sample period 2001–2020 using U.S. equity data with 61 firm characteristics. Sharpe ratios are annualized. Alphas are monthly returns (%) with respect to various factor models: CAPM, Fama-French 5-factor (FF5), Q5, RP-PCA 5-factor (RP5), and IPCA 5-factor (IP5). All alphas are statistically significant at the 1% level (*** $p < 0.01$). P-Tree1 uses a single tree with 10 leaf portfolios. P-Tree1-5 through P-Tree1-20 use boosted models with 5, 10, 15, and 20 trees respectively (50, 100, 150, 200 total leaf portfolios).

Table 12: Scenario C: OOS P-Tree Performance (Cong et al. 2025)

Model	Sharpe Ratio	α_{CAPM}	α_{FF5}	α_{Q5}	α_{RP5}	α_{IP5}
P-Tree1	4.35	1.50***	1.42***	1.35***	1.60***	1.58***
P-Tree1-5	3.87	1.18***	1.05***	0.96***	1.23***	1.24***
P-Tree1-10	4.29	1.02***	0.93***	0.85***	1.14***	1.10***
P-Tree1-15	4.03	0.96***	0.86***	0.80***	1.07***	1.02***
P-Tree1-20	3.88	0.96***	0.87***	0.81***	1.08***	1.03***

Source: Cong et al. (2025), Table 7, Panel C2. Results are for out-of-sample period 1981–2000 using U.S. equity data with 61 firm characteristics. Models are trained on 2001–2020 and tested on 1981–2000 (reverse time split). Sharpe ratios are annualized. Alphas are monthly returns (%) with respect to various factor models: CAPM, Fama-French 5-factor (FF5), Q5, RP-PCA 5-factor (RP5), and IPCA 5-factor (IP5). All alphas are statistically significant at the 1% level (*** $p < 0.01$). P-Tree1 uses a single tree with 10 leaf portfolios. P-Tree1-5 through P-Tree1-20 use boosted models with 5, 10, 15, and 20 trees respectively (50, 100, 150, 200 total leaf portfolios).