FINANCE PROJECT

To: Zeigham Khokher

By:
Poorvi Phulwani
Hanyin Yang
Gouri Maddulapalli
Irene Mugeni
Michelle Shih

Part 1: Backtesting Timing Strategies – Maximum Drawdown Analysis

This study evaluates three timing strategies—tbl_change, HML_rolling_12m, and bm_percentile—along with a combined strategy, to assess their effectiveness in reducing risk and improving returns compared to a static strategy, using maximum drawdown and Sharpe ratio as key metrics.

Backtesting Methodology: Using a 5-year rolling window OLS regression, the strategy invests in the market if a factor's coefficient is positive; otherwise, it moves to cash. Performance is compared to a fully invested static strategy.

Results Analysis:

	T-statistic	P-value	Beta Significance	Sharpe Ratio (Timing)	Sharpe Ratio (Static)	Max Drawdown (Timing)	Max Drawdown (Static)	Kolmogorov-Smirnov Statistic	Kolmogorov-Smirnov P-value
tbl_change	−1.519788	0.128825	0.072488	−1.367394	−1.366718	0.327234	0.417261	0.355848	1.608080e−34
HML_rolling_12m	0.002279	0.998182	0.662273	−1.366707	−1.366718	0.417261	0.417261	0.001647	1.000000e+00
bm_percentile	−0.867289	0.385955	0.161450	−1.370018	−1.366718	0.294826	0.417261	0.263591	6.056508e−19
combined	0.002279	0.998182	0.611203	−1.366707	−1.366718	0.417261	0.417261	0.001647	1.000000e+00

(Refer to Exhibit 1)

The bm_percentile strategy reduces drawdown but does not enhance risk-adjusted returns. Other strategies, like tbl_change and HML_rolling_12m, either show no improvement or slightly reduce performance. Statistical tests indicate weak predictive power, with only tbl_change and bm_percentile affecting return distributions. Overall, none of the strategies offer significant advantages over the static approach.

Part 2: Testing ML Models for Out-of-Sample R²

We test Logistic Regression, Random Forest, and XGBoost to predict if the HML factor will exceed its historical median, evaluating their impact on out-of-sample R^2.

ML Model 1 - Logistic Regression: *(Refer to Exhibit 2)*

- Features (X): tbl_change, HML_rolling_12m, bm_percentile, tbl_change_lag1 (scaled).
- Target (y): Binary (1 if next month's HML return is above median, else 0).

We back-test a 60-month rolling logistic regression model, retraining monthly to predict whether HML will exceed its median next month. The signal generation assigns a "buy" (1) if the predicted probability exceeds a threshold, otherwise "no-buy" (0). If invested, returns match HML's return; otherwise, returns are 0. Various thresholds are tested to optimize the Sharpe ratio. Trading costs of 10bps, 20bps, and 50bps are applied to simulate entry/exit expenses.

Results Analysis:

Bps Cost	Sharpe Ratio	T-stat	P-value
10	0.15	0.328	0.743
20	0.14	0.204	0.838
50	0.12	-0.167	0.868

The logistic regression strategy performs similarly to the Static HML Strategy at low trading costs, offering no significant advantage. However, as costs increase, the strategy's performance deteriorates further, making it less effective. The logistic regression strategy yields a modest Sharpe ratio of 0.159, offering no clear advantage over static strategies. Its out-of-sample R^2 (0.0009) indicates minimal predictive power in factor timing returns. While it may show slight benefits under low trading costs, its performance declines rapidly as costs rise. Low Sharpe ratios and insignificant t-tests suggest it is not superior to a static HML investment.

ML Model 2 - Random Forest: *(Refer to Exhibit 3)*

Using the same approach, bm_percentile, HML_rolling_12m, and tbl_change are employed as features to predict whether HML will exceed its median return next month.

Back-testing results:
ROC AUC Score: 0.652 – indicates moderate performance in distinguishing between the two classes (outperform HML or not).
Out-of-Sample R^2: 0.074 – the model explains only 7.4% of the variability in the target variable.
Sharpe Ratio: 0.38 – reflects a moderate risk-adjusted return.
Max Drawdown: 0.26 – the maximum loss is 26%.

In conclusion, the random forest strategy has limited explanatory power but outperforms the Static HML Strategy in cumulative returns, capturing periods of outperformance through dynamic HML exposure adjustment.

ML Model 3 - XGBoost: *(Refer to Exhibit 4)*

We adopt a similar approach as the previous models, using bm_percentile, HML_rolling_12m, and tbl_change as features for the Gradient Boosting model to predict whether the HML factor return in the next month will be above (1) or below (0) its median return.

Back-Testing Results Analysis:
ROC AUC Score: 0.546 – modest ability to distinguish HML outperforming or underperforming its historical median.
Out-of-Sample R^2: -0.022 – poor explanatory power for variability in the target variable.
Sharpe Ratio: 0.37 – slight improvement over the Random Forest strategy, indicating better risk-adjusted return.

Max Drawdown: 0.159 – a 15.9% maximum loss, suggesting better risk management compared to Random Forest.

The Gradient Boosting strategy improves risk management with a higher Sharpe ratio and lower drawdown but lacks strong explanatory power. It captures HML outperformance well but doesn't significantly outperform a static strategy. Key predictors include HML_rolling_12m and bm_percentile, highlighting momentum and value factors. The cumulative returns plot shows the Gradient Boosting Timing Strategy (red line) outperforms static HML (gray dashed line) in the long run, similar to Random Forest. However, improvements are modest, indicating the need for further tuning.

The results show that while machine learning models offer slight improvements, their out-of-sample performance is limited. Logistic regression had minimal predictive power, Random Forest showed slight improvement, and Gradient Boosting exhibited overfitting. Despite better Sharpe ratios and lower drawdowns, the improvements were too small to surpass a static strategy.

Part 3: Managing Trading Costs in Factor-Based Strategies

Trading costs significantly impact the performance of factor-based investment strategies. In this section, we evaluate the effect of proportional transaction costs, analyze portfolio turnover at different rebalancing frequencies, and explore quadratic transaction costs (market impact model) to determine optimal cost management strategies. Our goal is to balance return maximization with trading cost minimization while maintaining effective factor exposure.

Step 1:Proportional Transaction Costs - Impact on Factor Returns (Refer to Exhibit 5)

To assess the effect of trading costs, we apply proportional transaction costs of 10bps, 20bps, and 50bps to our factor-based strategy.

Key Findings:

- *10bps:* Minimal impact, cumulative return remains close to no-cost scenario.
- *20bps:* Noticeable decline in net returns.
- *50bps:* Significant reduction in returns; strategy sustainability is questionable.

The line graph of cumulative returns shows a progressive decline in performance as transaction costs increase. The 50bps cost line deviates sharply from the no-cost scenario, while the 10bps and 20bps lines remain closer but still exhibit lower final values. Even a small increase in trading costs compounds over time, reducing overall returns. High turnover strategies are particularly vulnerable, making cost-efficient execution crucial for preserving profitability.

Step 2: Portfolio Turnover & Rebalancing Frequency (Refer to Exhibit 6)

Portfolio rebalancing frequency determines the number of trades, directly influencing transaction costs. We analyze monthly, quarterly, and semi-annual rebalancing to compare cost implications.

Key Findings:

- *Monthly rebalancing:* Highest turnover, resulting in frequent and costly trades.
- *Quarterly rebalancing:* Moderately reduces turnover while maintaining timely factor adjustments.
- *Semi-annual rebalancing:* Lowest turnover, significantly minimizing transaction costs.

The line graph of portfolio turnover reveals frequent trading spikes in the monthly strategy, highlighting excessive costs. In contrast, quarterly and semi-annual turnover lines remain stable and lower, indicating less trading activity.

Reducing rebalancing frequency minimizes trading costs while maintaining reasonable factor exposure. Quarterly or semi-annual schedules strike the best balance between cost efficiency and responsiveness to market trends.

Step 3: Quadratic Transaction Costs (Market Impact) (Refer to Exhibit 7)

As capital investment increases, market impact costs grow non-linearly. We model quadratic transaction costs at different capital levels ($1M, $5M, $10M, and $50M) to determine at what point factor premiums vanish.

Key Findings:

- *$1M - $5M:* Market impact costs are manageable; strategy remains profitable.
- *$10M - $50M:* Costs scale significantly, reducing net returns, with $50M experiencing the sharpest decline.

The line graph of net returns across different capital levels confirms that higher capital levels face greater cost pressure, leading to more volatile and declining returns. The $1M capital strategy remains stable, while the $50M strategy experiences significant performance deterioration.

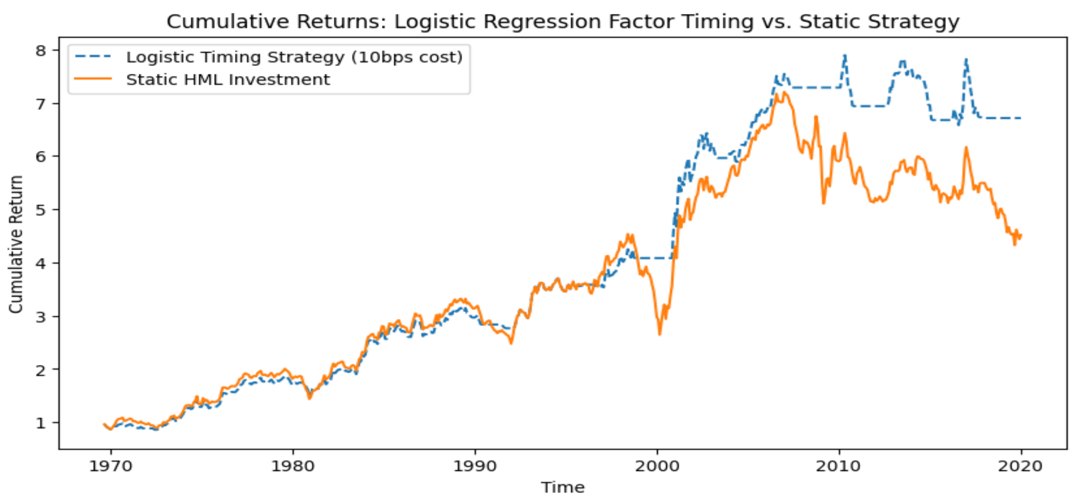
Factor strategies are highly sensitive to capital size, with market impact costs eroding profitability beyond $10M-$50M. Large-scale investors must employ cost-efficient execution methods, such as algorithmic trading, to mitigate cost pressures.

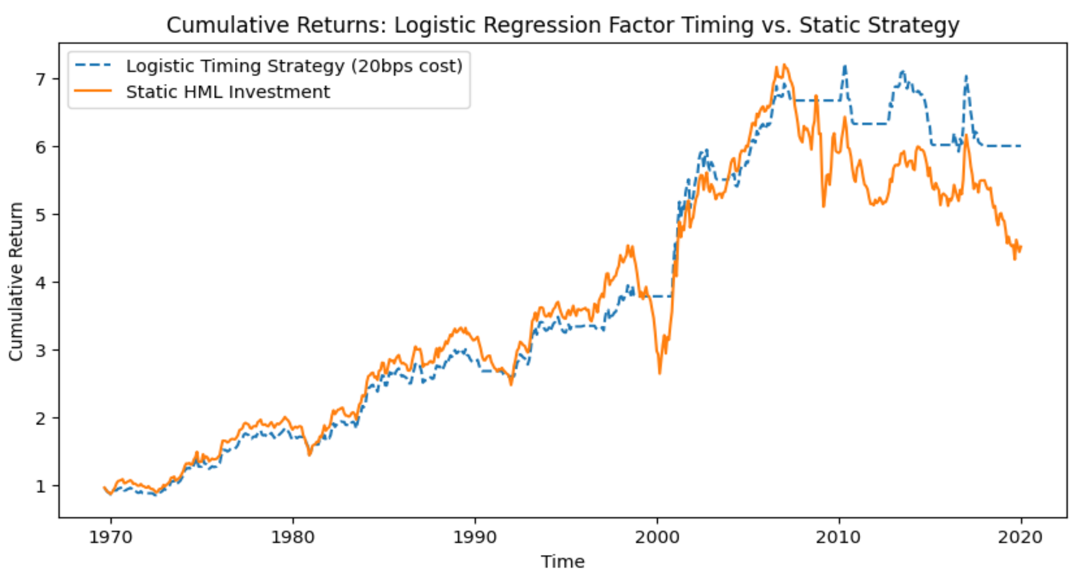
Exhibits:

Exhibit 1: Backtesting Timing Strategies – Maximum Drawdown Analysis

```
                           Logit Regression Results
==============================================================================
Dep. Variable:                    HML   No. Observations:                  666
Model:                          Logit   Df Residuals:                      661
Method:                           MLE   Df Model:                            4
Date:                Fri, 28 Feb 2025   Pseudo R-squ.:                 0.01753
Time:                        01:45:45   Log-Likelihood:                -453.54
converged:                       True   LL-Null:                       -461.63
Covariance Type:            nonrobust   LLR p-value:                  0.002781
===================================================================================
                      coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------
const              -0.0064      0.078     -0.081      0.935      -0.160       0.147
tbl_change          0.1147      0.088      1.307      0.191      -0.057       0.287
HML_rolling_12m     0.2449      0.082      2.997      0.003       0.085       0.405
bm_percentile       0.1182      0.080      1.481      0.139      -0.038       0.275
tbl_change_lag1     0.0988      0.086      1.143      0.253      -0.071       0.268
===================================================================================
```

Exhibit 2: ML Model 1 - Logistic Regression





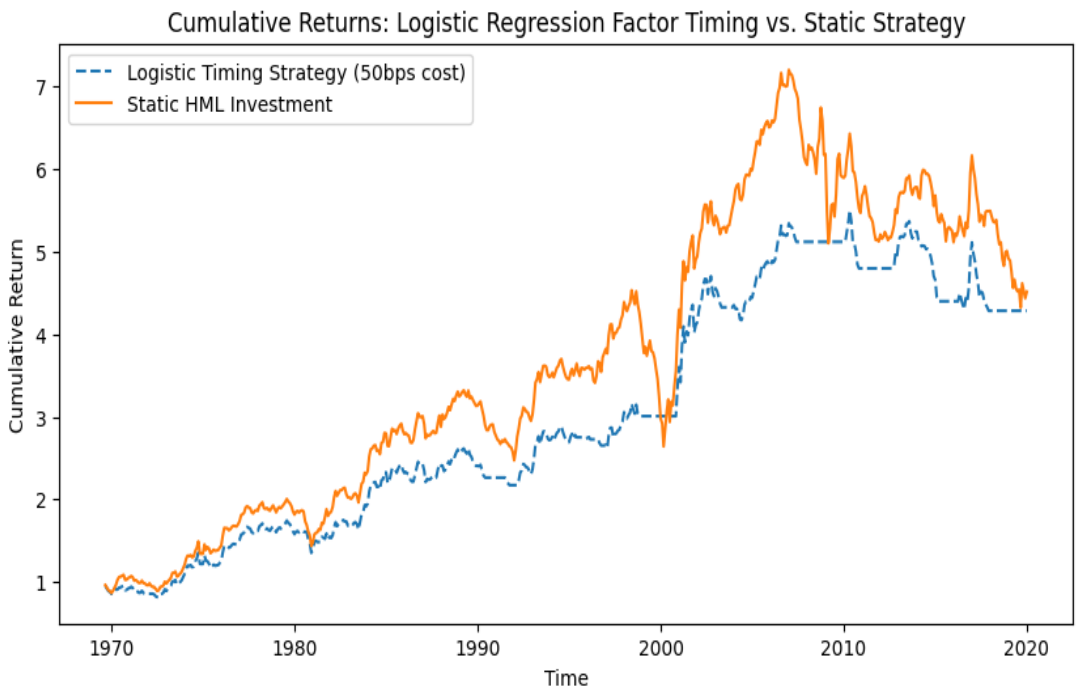
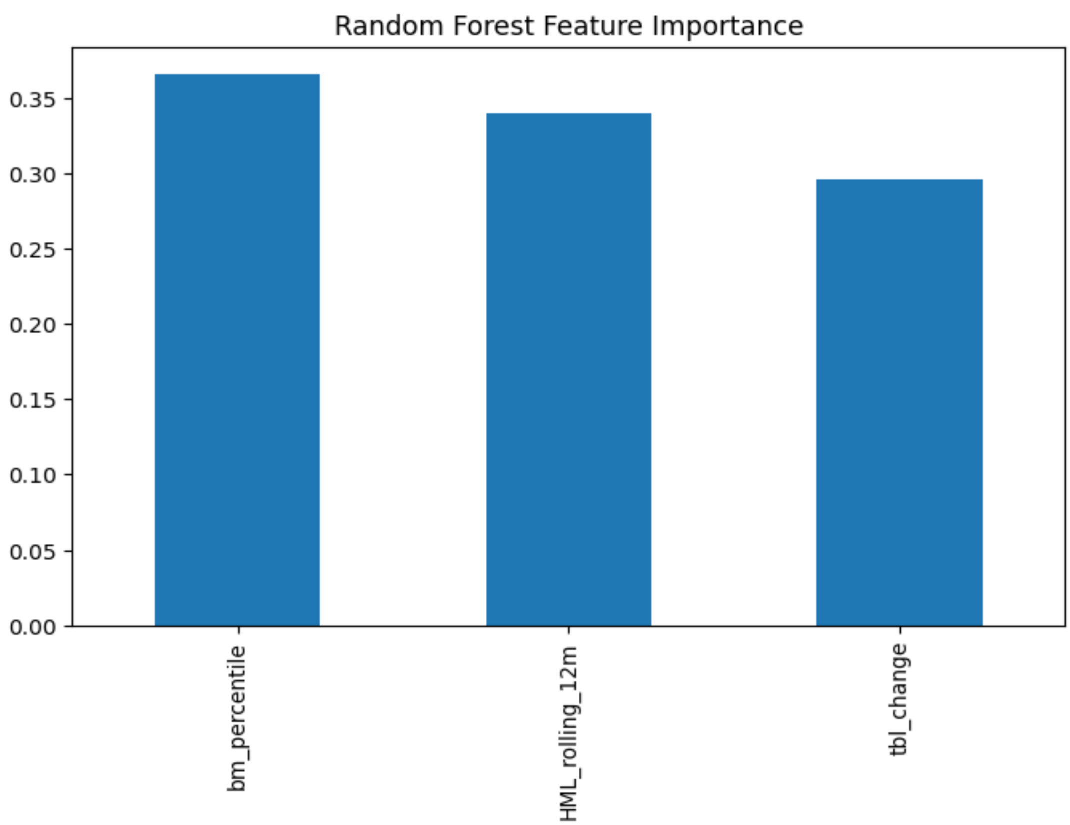
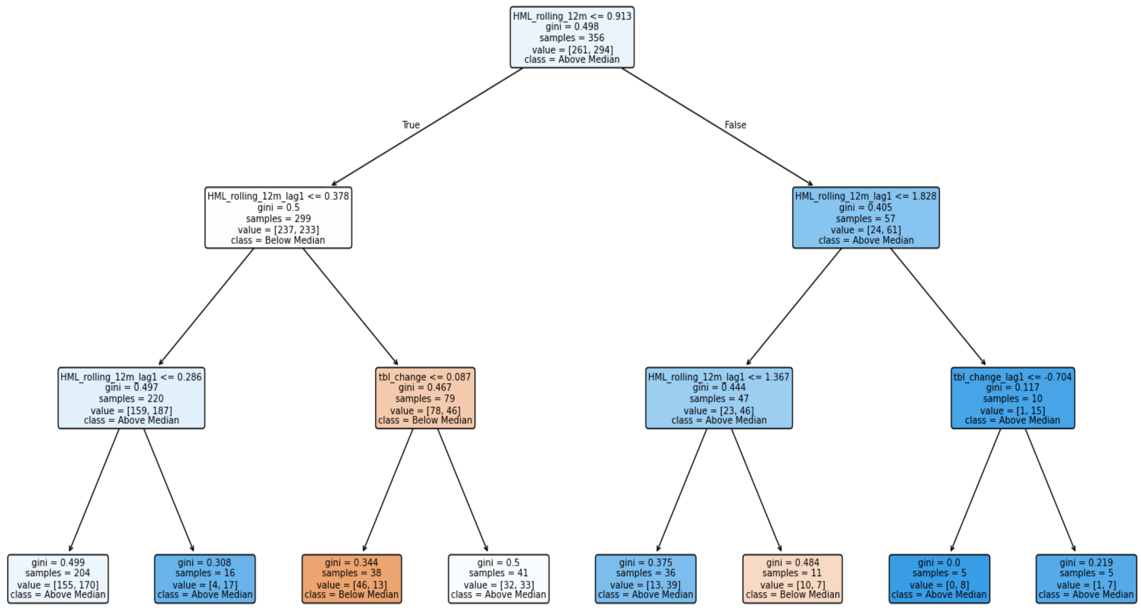


Exhibit 3: ML Model 2 - Random Forest



Visualization of a Single Decision Tree



Random Forest Timing Strategy vs. Static Strategy

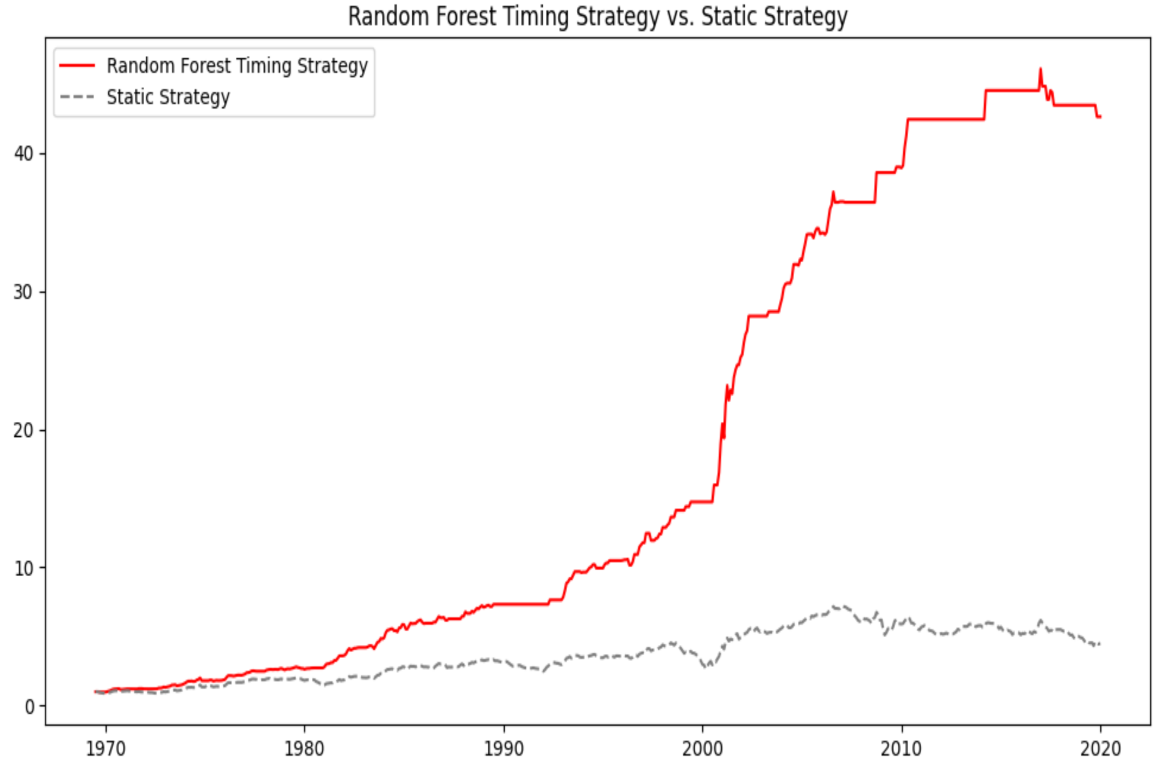
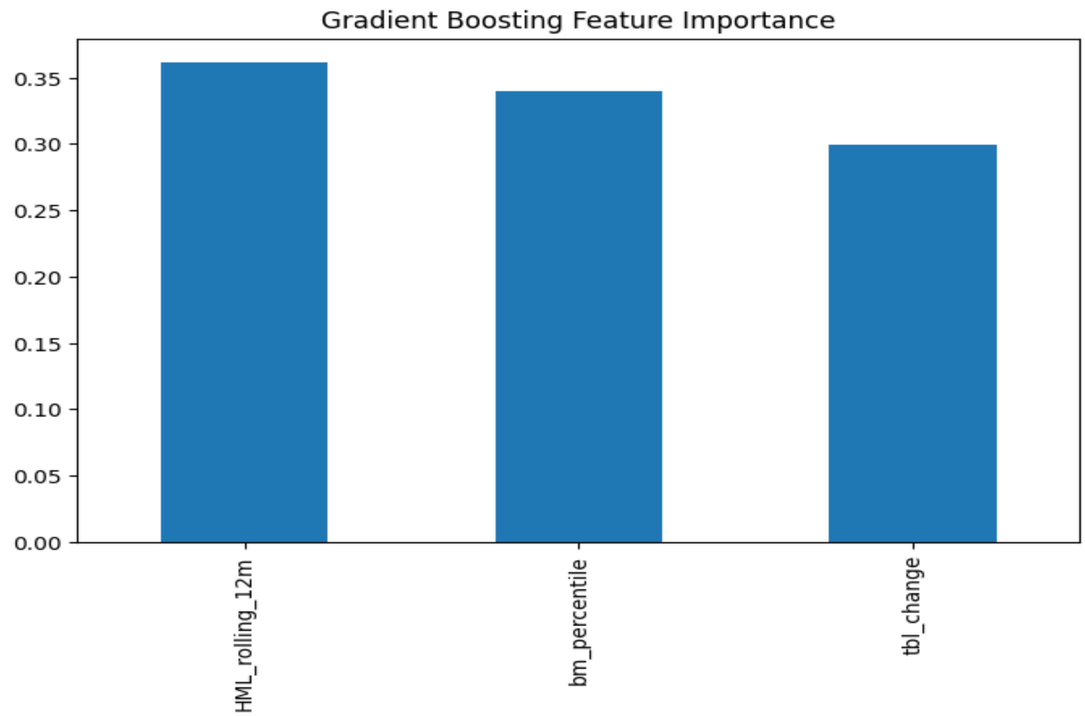


Exhibit 4: ML Model 3 - XGBoost



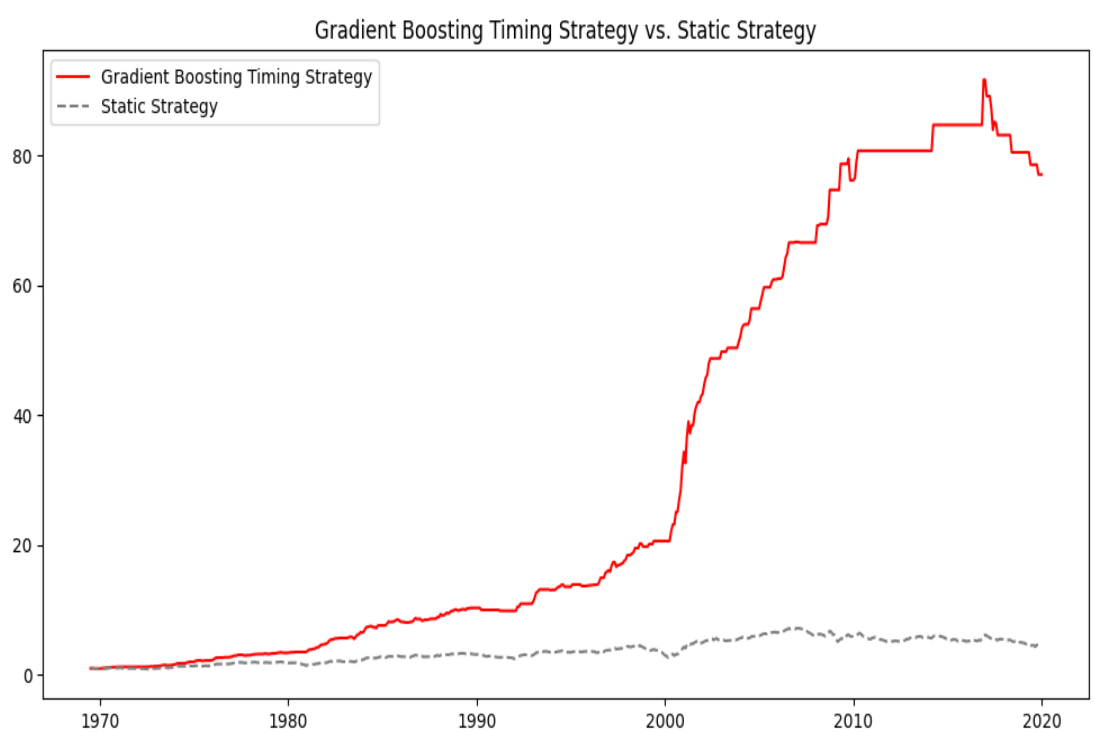


Exhibit 5: Proportional Transaction Costs - Impact on Factor Returns

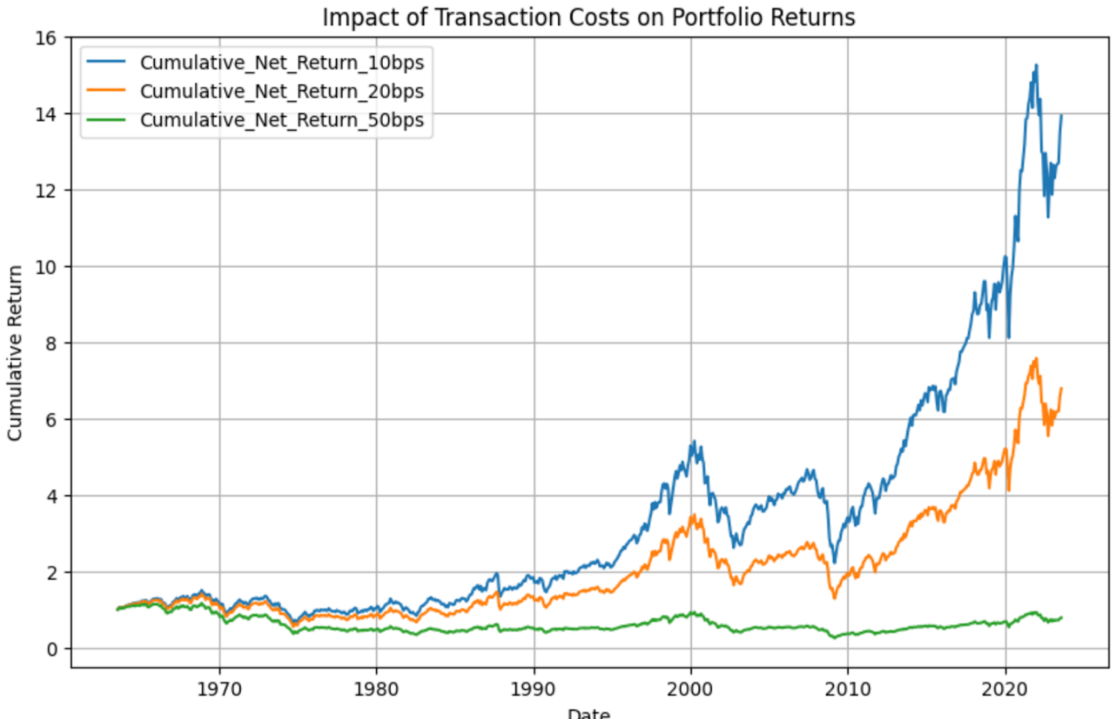


Exhibit 6: Portfolio Turnover & Rebalancing Frequency

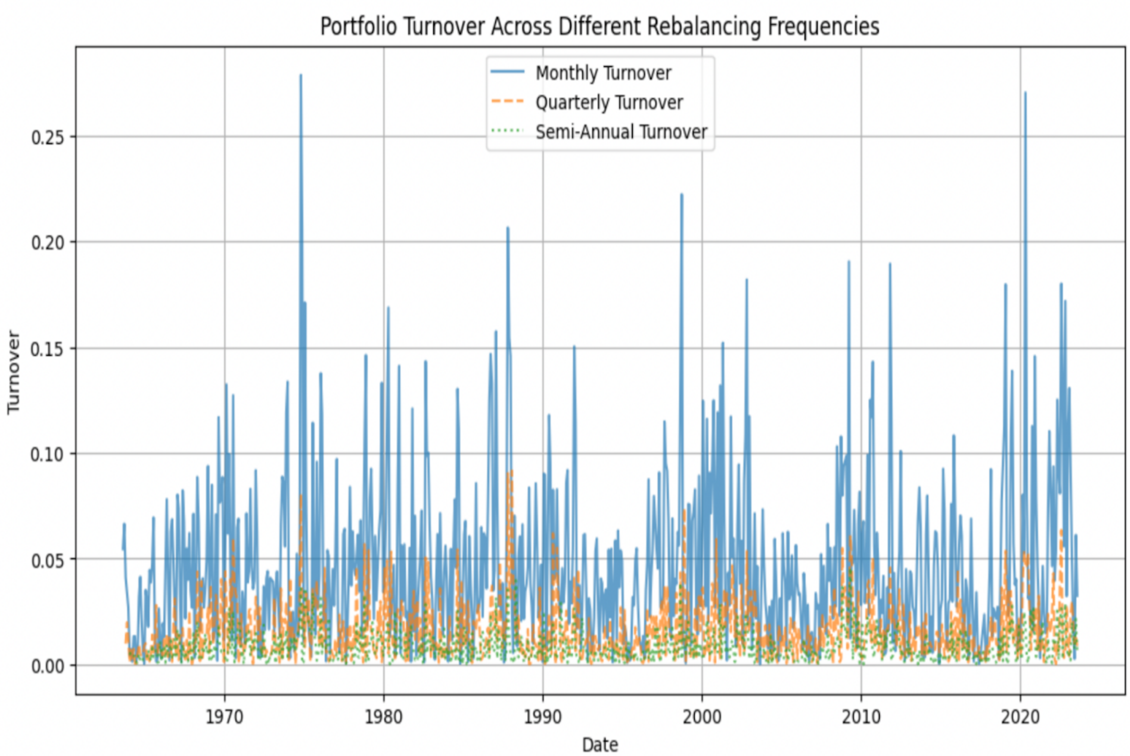


Exhibit 7: Quadratic Transaction Costs (Market Impact)