3. Karanaugh M.C., Trussell R.R. Desing of aeration towers to strip volatile contaminats from drinking water//Journal of American Water Works Association. – 2010. – 72, N12. – pp.684-692.

4. Kozhinov V.Ch., Kozhinov I.V. Ozonation of water. M: Stroyizdat, 1973, – p.160s.

УДК 336.76
ББК 65.05

LONG-SHORT EQUITY PORTFOLIOS: PERFORMANCE EVALUATION OF MULTI-DIMENSIAL STRATEGIES

Nadzeya Zarubina, FRM
MSc in Quantitative Finance, Vienna University of Economics and Business, Vienna, Austria
Raiffeisen Bank International, Financial Planning and Analysis team, Vienna, Austria

The main objective of the research is to estimate the backtested performance of multi-dimensional equity long-short strategies, which were constructed based on a combination of different signals (fundamental indicators). An evaluation of performance is conducted using the appropriate t-tests (derived by Novy-Marx, "Backtesting Strategies Based on Multiple Signals" [2015]) by estimating the statistical significance of the backtested average weekly returns for both the EU and US markets. The data set includes weekly stock prices of 2 553 firms for the period January 1990 to November 2015 for the US market and January 2000 to November 2015 for the EU market. The obtained results show that the combinations of signals provide statistically significant results for 1 out of 48 portfolios (under the assumption of zero transactions costs).

Keywords: backtested performance, data mining, overfitting bias, selection bias, multiple signals, stock returns, long-short strategy, fundamental indicators.

INTRODUCTION

Buoyed by exceptional economic and business conditions, returns on US and Western European equities during the past 30 years were considerably higher than the long-run trend, says McKinsey Global Institute in their report: "Diminishing returns: why investors may need to lower their expectations" [May, 2016].

Over the past years market returns have become extremely volatile. As a consequence of this trend, most investors make emotionally-based decisions that focus on the near-term future and ignore longer-term opportunities.

To make the best investment plans for the future, investors need access to unbiased, long-term performance results. In this regard, it is important for an investigator to be aware of the existence of biases in the backtested performance results. Particularly, multi-dimensional strategies, based on a combination of multiple signals suffer from severe selection and overfitting biases. Selection bias results when the investigator considers more signals than he employs. While overfitting bias arises when each signal is used so that it individually predicts positive in-sample returns. Hence, performance of multi-dimensional strategies should be evaluated differently, by accounting for those biases. This issue is addressed by Novy-Marx, «Backtesting Strategies Based on Multiple Signals» [2015].

The research question of this article is the following: How significant are the returns of US and EU multi-dimensional long-short equity strategies in the presence of selection and over-fitting biases?

In the course of the research it will be analyzed whether it is possible to build equity portfolios based on a combination of fundamental indicators (signals) which have statistically and economically significant average weekly returns in the presence of selection and overfitting biases.

The research carried out in this article is done under the following hypothesis:

Hypothesis: Employing a combination of signals in construction of a strategy leads to positive returns.

The literature that I refer to in this article can be divided into 3 groups:

- the first group describes the studies of stock selection criteria and their relationship to the abnormal returns;

- the second group demonstrates examples of the currently famous and widely used multi-signal strategies which are used in investment practices. Those strategies rely on a composite measure that combines multiple signals;

- the third group is formed by papers that analyze potential dangers of data mining and multiple testing. Those papers are also summarizing the implications and risks of: using backtested performance of trading strategies as an indicator of effectiveness and profitability; implementation of the strategy in real world.

EMPIRICAL RESULTS

1. Data description

The data set includes weekly returns of equities that are traded in the US and EU markets. The timeframe covered by the sample is 25 years for the US market (January 1, 1990 - November 13, 2015) and 15 years for the EU market (January 1, 2000 - November 13, 2015). The total number of companies under the consideration is 2 553 (1 437 for the EU market and 1 116 for the US market). A firm's stock market performance is often evaluated in comparison to a benchmark: the industry or the whole economy.

2. Strategy construction

This section describes the approaches used to build long-short equity multi-dimensional strategies. The long-short strategy used in this research implies that 100% of the amount is invested in a risk-free security, 100% long in 30 companies with the highest score and 100% short in 30 companies with the lowest score:

100% risk-free + 100% the highest - 100% the lowest

In this way, since we have equal amounts of investment in both long and short positions, net market exposure is completely eliminated under the assumption that risk-free return is equal to zero.

To separate "the highest score companies" from "the lowest score companies" in this research, the following fundamental indicators were employed:

1. BTP: book-to-price ratio (also called the book-to-market ratio) is used as a measure of relative value of equity.

2. Beta1Y: is an indicator used to measure the market risk of equity.

3. Corporate Finance: is a fundamental indicator which was calculated as a compound score and includes the following factors: post-earnings announcement drift, neo classical factor, share-buyback (negative change in number of outstanding shares) and dividend omissions / initiations.

4. DY: Dividend Yield. Calculated as a dividend per share (paid in the last year) divided by the current price of the share.

5. Leverage: calculated as a debt divided by a common equity. Usually the leverage ratio is very high for banks.

6. Momentum: is the 1 year-return of a stock shifted by one month. Calculated as: $\frac{R_{t-4}}{R_{t-56}} - 1$; where the lags are given in weeks.

7. PE: price-to-earnings ratio is a firm's stock price divided by earning per share.

8. Quality: is a compound score of long-term return on assets and high earnings predictability.

9. Size: is the market capitalization of a company in portfolio currency (USD or EUR) and represents a company's equity value on the stock market. This indicator is calculated as a number of shares outstanding multiplied by a price per share.

In this research the following 3 different methods were used to construct strategies based on a combination of signals:

1. by combining the strongest signals for the EU market;
2. by combining the strongest signals for the US market;
3. by combining economically meaningful signals (according to the current academic research).

Method 1: EU signals. This method is based on the idea that we combine the strongest signals applicable to the European equity market. Table 2.1 shows an overview of the portfolios annualized HPR (Holding Period Return) based on the pure signal.

Table 2.1 — Portfolios annualized HPR for 9 one-dimensional long-short strategies for the EU market. Shows a potential informative power of a signal employed in a strategy construction; sorted by the absolute values of returns in a descending order.

Pf Returns, pure signal. EU

	Ret	Ret.ABS
PE	-0.158	0.158
Beta	-0.106	0.106
Size	-0.083	0.083
DY	0.082	0.082
Leverage	-0.048	0.048
Corporate Finance	0.046	0.046
Momentum	0.034	0.034
BTP	0.011	0.011
Quality	0.004	0.004

Based on the presented results above, we can conclude that the strategy based on the PE score is giving the highest return in absolute terms: 15,8%.

Method 2: US signals. This method is similar to the Method 1 with the only exception that we backtest the portfolios comprising the stocks of the American companies to find the most informative signals applicable to the US equity market.

Table 2.2 — Portfolios annualized HPR for 9 one-dimensional long-short strategies for the US market. Shows a potential informative power of a signal employed in a strategy construction; sorted by the absolute values of returns in a descending order.

Pf Returns, pure signals. US

	Ret	Ret.ABS
Size	-0.101	0.101
DY	-0.092	0.092
PE	-0.066	0.066
Quality	-0.047	0.047
Momentum	-0.031	0.031
Beta	-0.017	0.017
Corporate Finance	0.014	0.014
Leverage	-0.008	0.008
BTP	0.002	0.002

Analyzing and comparing the results of backtesting in the US and EU markets, we can observe that the signals which work well for one market, do not necessarily work well for another market

Method 3: Academic signals. This method implies that the signals were combined based not on informative power but on findings of the current academic research (some combinations of signals proved to be profitable).

3. Limitations of backtesting

Backtesting - is the application of a quantitative model to historical market data to generate hypothetical performance during a prior period. The main goal of backtesting is to show performance returns that would have been achieved if the investment approach had been in existence during the tested period and before the real capital is invested.

In particular, when investigator considers signals and combines the best *k* of those to select stocks, strategies may suffer from:

1. Pure selection bias (or multiple testing bias): results when the researcher considers more signals than he employs.

2. Pure overfitting bias: results in case the researcher uses all the signals considered and underlying signals are typically signed in such a way that each predicts positive in-sample returns.

3. Combination of selection and overfitting biases: while the overfitting and selection biases are distinct, they do interact, with the selection bias severely exacerbating the overfitting bias.

4. Theoretical model: critical value approximation

The aim of this section is to show how the theoretical distributions for critical t-statistics were derived by NM in "Backtesting Strategies Based on Multiple Signals" [2015]. The results reported in academic research with respect to trading strategies often suffer from the issue of data mining.

Table presented below shows computed adjusted critical t-values that should be taken into account when hypothesis testing of multi-dimensional strategies for significance is conducted.

Table 3.1 — Theoretical Model 5% critical t-values. This table shows the critical t-values adjusted for selection and overfitting biases for the best *k-of-n* strategies on a 5% significance level.

Strategy	Critical t-value
General cases	
best 2 - of - 9	3.217
best 3 - of - 9	3.472
best 4 - of - 9	3.609
best 5 - of - 9	3.668
best 6 - of - 9	3.670
best 7 - of - 9	3.626
best 8 - of - 9	3.540
Special cases	
best 1 - of - 9 (pure selection bias)	2.736
best 9 - of - 9 (pure overfitting bias)	3.416

From the table above we can observe a following pattern: in general, the more signals we combine to build a strategy, the more critical t-statistics we obtain.

5. Performance evaluation

This section presents the components determining a portfolio's investment performance: cumulative returns, average weekly returns, average weekly excess returns, the volatility of returns and Sharpe Ratio for 24 long-short equity strategies defined in the previous chapter.

Figure 4.1 presented below shows the backtested performance of the TOP 5 Portfolios in comparison with the performance of the benchmark in the EU market.

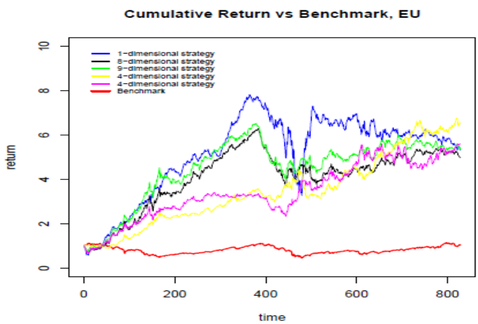

Figure 4.1 — Backtested Performance, TOP 5, EU. This plot shows the backtested performance of the TOP 5 Portfolios and the Benchmark for the EU market; January 1, 2000 - November 13, 2015.

Similarly, the same impressive results relative to the benchmark could be obtained if an investor follows:

• PE inversion, Beta inversion, Size inversion, DY, Leverage inversion, Corporate Finance, Momentum and BTP strategy (the black line)

• PE inversion, Beta inversion, Size inversion, DY. Leverage inversion, Corporate Finance, MoM, BTP and Quality strategy (the green line)

• Beta inversion, PE inversion, Size inversion and Leverage inversion strategy (the purple line).

The backtested performance of the TOP 5 portfolios and a benchmark for the US market is presented on the graph below:

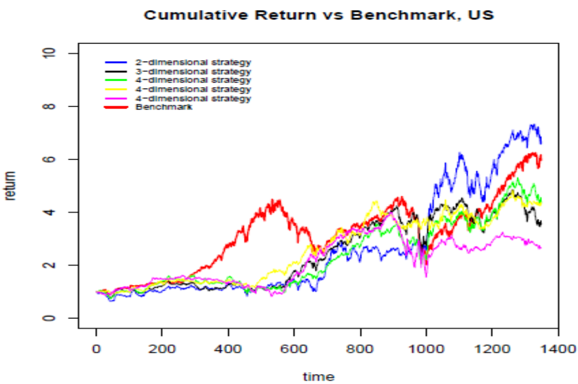

Figure 4.2 — Backtested Performance, TOP 5, US. This plot shows the backtested performance of the TOP 5 Portfolios and the Benchmark for the US market; January 1, 1990 - November 13, 2015.

HYPOTHESISTESTING RESULTS. GRAPHICAL REPRESENTATION

This section is a central part of the Research. It summarizes and illustrates the statistical significance of constructed multi-dimensional long-short portfolios by presenting the figures with the critical and estimated t-statistics.

Figure 4.3 illustrates the distribution of the adjusted for selection and overfitting biases critical t-values and the estimated t-statistics for the EU and US markets when we construct portfolios on the EU and academic signals.

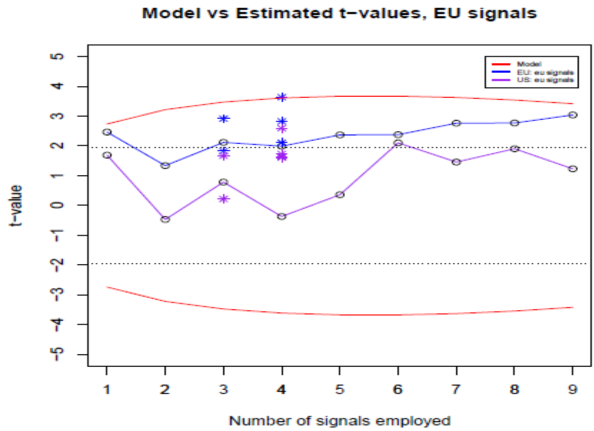

Figure 4.3 — Model critical t-values (solid red line) and the estimated t-statistics of the average weekly returns of the portfolios constructed by combining the best EU signals (solid lines) and the academic signals (stars) for the EU and US markets.

Figure 4.4 presented below illustrates the standard critical t-values, the adjusted critical t-values and the calculated t-statistics of the average weekly returns of the portfolios based on the US and academic signals.

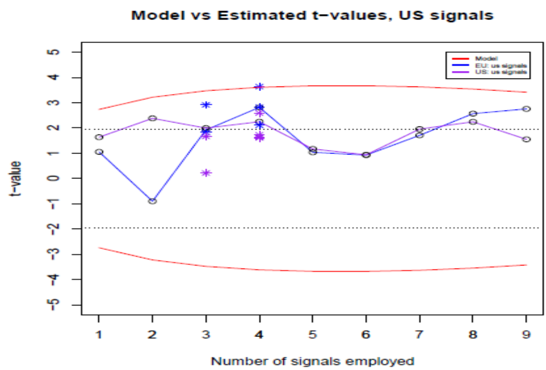

Figure 4.4 — Model critical t-values (solid red line) and the estimated t-statistics of the average weekly returns of the portfolios constructed by combining the best US signals (solid line) and the academic signals (stars) for the EU and US markets.

The following 2 graphs distinguish the hypothesis testing results between the EU and US markets. Figure 4.5 illustrates the standard critical t-values, the adjusted critical t-values and the computed t-statistics of returns in the EU stock market only. Figure 4.6 shows the theoretical t-values and the estimated t-values for the US market.

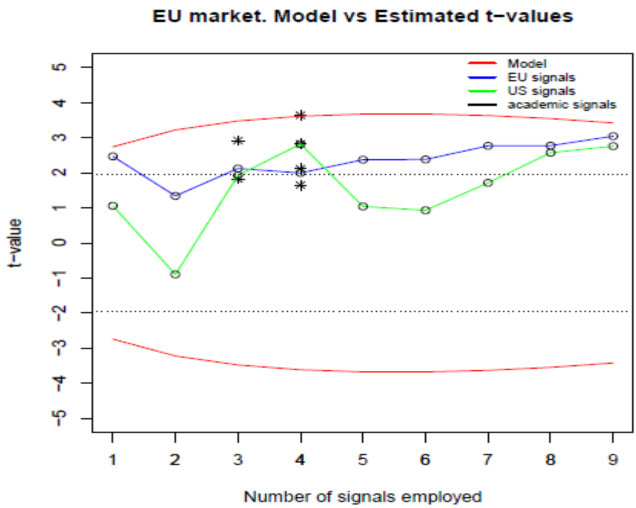

Figure 4.5 — Model critical t-values (solid red line) and the estimated t-statistics of the average weekly returns of the portfolios constructed by combining the best EU signals (blue solid lines), the best US signals (green solid lines) and the academic signals (black stars) in the EU market.

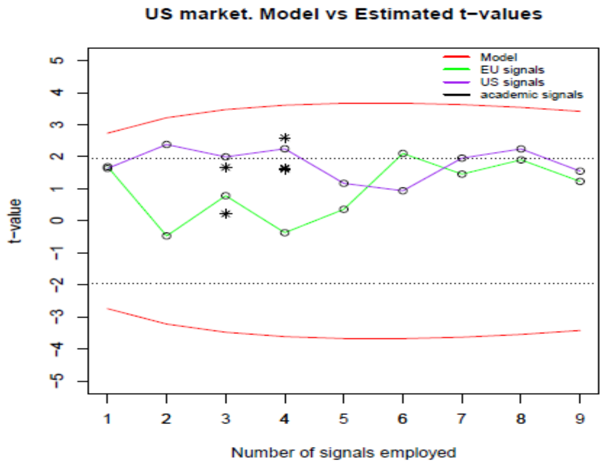

Figure 4.6 — Model critical t-values (solid red line) and the estimated t-statistics of the average weekly returns of the portfolios constructed by combining the best EU signals (green solid lines), the best US signals (purple solid lines) and the academic signals (black stars) in the US market.

This Research highlights one of the effects of data mining that investors may or may not be aware of. Because of data mining, many potentially profitable investments strategies that work in backtesting do not perform as well when implemented. Theoretical model of critical t-values approximation derived by NM has a substantial contribution to finance and investments area.

CONCLUSION

Based on theoretical and empirical research and taken into account described assumptions, it can be concluded that the hypothesis of the existence of different from zero weekly returns of multi-dimensional long-short strategies among the SP 500 and JD STOXX constituents for the specified periods: January 1990 - November 2015 (SP 500) and January 2000 - November 2015 (DJ STOXX) must be rejected using appropriate t-tests in 47 cases out of 48. It is important to distinguish between a statistical result and an economically meaningful result, because statistical significance does not necessarily imply economic significance due to transactions costs, taxes and risk preferences. It is highly probable that removing zero transaction cost assumption from our research will lead to economically meaningless results.

REFERENCES

1. Asness, Clifford S., Frazzini, Andrea, and Pedersen, Lasse Heje. “Quality Minus Junk". In: (2015). doi: 10.2139/ssrn.2312432.
2. Asness, Clifford S., Moskowitz, Tobias J., and Pedersen, Lasse Heje. “Value and Momentum Everywhere". In: The Journal of Finance 68.3 (2013), pp. 929{985. issn: 1540-6261. doi: 10.1111/jofi.12021. URL: http://dx.doi.org/10.1111/jofi.12021.
3. Baker, Malcolm and Wurgler, Jeffrey. “Investor Sentiment and the Cross-Section of Stock Returns". In: Working Paper Series 10449 (2004). doi: 10.3386/w10449. URL: http://www.nber.org/papers/w10449.
4. Bali, Turan G., Cakici, Nusret, and Tang, Yi. “The Conditional Beta and the Cross-Section of Expected Returns". In: Financial Management 38.1 (2009), pp. 103{137. URL: http://EconPapers.repec.org/RePEc:bla:finmgt:v:38:y:2009:i:1:p:103-137.
5. Blake, Christopher R, Elton, Edwin J, and Gruber, Martin J. “The Performance of Bond Mutual Funds". In: The Journal of Business 66.3 (1993), pp. 370{403. URL: http://EconPapers. repec.org/RePEc:ucp:jnlbus:v: 66:y:1993:i:3:p:370-403.
6. Dudler, Martin, Gmuer, Bruno, and Malamud, Semyon. “Risk Adjusted Time Series Momentum". In: Social Science Research Network Working Paper Series (2014). URL: http://ssrn.com/abstract=2457647.
7. Fama, Eugene F. and French, Kenneth R. “Size, value, and momentum in international stock returns". In: Journal of Financial Economics 105.3 (2012), pp. 457{472. URL: https://ideas.repec.org/a/eee/jfinec/v105y2012i3p457-472.html.
8. Frazzini, Andrea and Pedersen, Lasse Heje. “Betting Against Beta". In: Working Paper Series 16601 (2010). doi: 10 . 3386 / w16601. URL: http://www.nber.org/papers/w16601.