APPLIED FINANCE LETTERS

VOLUME 11, 2022

ARE FACTOR INVESTING STRATEGIES SUCCESSFUL OUT-OF-SAMPLE: EVIDENCE FROM THE NIFTY INDICES

VAIBHAV LALWANI[1*]

Xavier School of Management, XLRI Delhi-NCR Campus, Haryana, India

* Corresponding Author: Vaibhav Lalwani, Assistant Professor (Finance), XLRI Delhi-NCR Campus Aurangpur Village, Dadri Toye, Untloda, Haryana 124103, India * vaibhavlalwani@outlook.com

Abstract

Do factor investment strategies that have generated superior returns in the past continue to do so out-of-sample? To test this hypothesis, I check the performance of nine factor-based indices of the National Stock Exchange (NSE) of India. My results show that the performance of most indices falls considerably in the out-of-sample period, i.e., the period after the launch of an index. The results hold for absolute as well as excess and risk-adjusted returns. In additional tests, I find that none of the factor strategies generates significant alpha after controlling for standard factors such as size, value, and momentum. The results are robust to the exclusion of the COVID-19 period.

Keywords: factor investing; anomalies; asset-pricing

1. Introduction and Literature review

Beginning with the seminal studies of Banz (1981), Fama and French (1993), and many others, factor investing has exploded in popularity in academia and the industry. Factor investing involves picking stocks based on certain metrics that are supposed to predict future returns. These metrics could be valuation ratios (such as earnings to price, book value to price, etc.) or other fundamental or technical indicators of a company's profitability and financial strength. Some strategies use one factor (single factor), whereas some use multiple metrics (multi-factor) to rank and filter stocks. As per Hou, Xue, and Zhang (2020), more than 400 predictors[1] have already been established in the asset pricing literature, thus lending support to those who believe that stock returns are at least partially predictable. Riding on the academic success of factor investing, the financial services industry has also responded by providing avenues for investors who want to put their money in factor-based funds. As per BlackRock's estimates, the amount invested in factor funds is expected to be around $3.4 trillion by 2022.

The popularity of factor investing is not surprising. These strategies provide a healthy compromise between active and passive investing. First, they condense the numerous potential signals used by active investors to a chosen handful whose utility is backed by historical performance, thus reducing substantial complexity from the stock-picking process. Second, they allow investors a chance to beat the market by systematically filtering assets that may be underpriced (or may provide higher returns in the future).

Factor investing may have its proponents, but it also has its share of critics. Many are still skeptical of the robustness of these strategies. It is still unclear whether the many factors discovered in the literature

[1] I use the words predictors, factors, and anomalies interchangeably.

result from genuine patterns or relentless data snooping. Further, researchers are still divided about the source of the superior performance, i.e., whether additional returns to factors are due to risk or irrational mispricing.

As a result, multiple studies have tried to check for the out-of-sample performance of factor investing strategies, i.e., whether they perform beyond the sample in which these strategies were first discovered. Most, if not all, factors were first discovered in the U.S. market. Later, different researchers tested whether these factors provided abnormal performance outside the U.S. and beyond the sample period used by initial studies. The goal here is not to review this vast literature. Instead, I survey some recent studies that test for out-of-sample performance of factor investing strategies and show how this paper materially differs from the existing literature.

Mclean and Pontiff (2016) is a highly influential study testing for the out-of-sample performance of factor anomalies. They show that anomaly performance declines by as much as 58% post-publication. Linnainmaa and Roberts (2018) similarly report that most factor portfolios' returns decline out-of-sample while their volatilities and cross-correlations increase. Hollstein (2022), on the other hand, shows that anomalies persist internationally in equally-weighted portfolios but largely disappear when excluding the impact of microcaps. Hou, Xue, and Zhang (2020) report similar findings in the U.S. market. Cakici et al. (2021) utilise hand-collected data from 1926-1987 and show that most anomalies do not replicate for stocks listed on the Stock Exchange of Melbourne. While many such studies report poor out-of-sample performance of factor investing, others suggest that factor-based strategies are still robust. For example, Jacobs and Müller (2020) show that the United States is the only country with a reliable post-publication decline in anomaly performance. They report robust performance of factor-based strategies in an international sample. Huang and Huang (2014) similarly find that anomalies persist out-of-sample, even after controlling for transaction costs.

Ultimately, the jury is still out on the superiority of factor investing. I contribute to this divided literature by testing for the out-of-sample performance of factor investing strategies by using a sample of factor strategy-based indices constructed by the NSE indices Ltd., i.e., a subsidiary of the largest stock exchange in India, the National Stock Exchange (NSE). I utilise NSE's Nifty strategy indices to compare the in-sample performance of factor-based strategies with their out-of-sample performance.

There are many indices in NSE's basket, and each one follows a different strategy for picking stocks. Investors can invest in products linked to these indices to meet their investment objectives. A typical index is released (or launched) to market participants after conducting a back-test from the base date up to the launch date. This back-test shows the strategy's performance from some pre-decided base date up to the launch date of the index. Therefore, the index's performance up to its launch is the in-sample or training data period performance. The index's performance after this period will be the out-of-sample or the test data performance.

My study differs from the usual factor investing literature by using index portfolios as test assets. The typical factor investing study involves the researcher herself sorting stocks into multiple buckets based on some indicators and creating portfolios of assets, and finally constructing long-short factors from these portfolios. As pointed out by Harvey (2017), this method offers too many degrees of freedom to the researcher and combined with a publication bias in favour of positive results, this broad methodology is likely to bias results in favour of the outperformance of a factor.

It becomes imperative to test whether factor-investing genuinely works out-of-sample because Lo and MacKinlay (1990) and Harvey (2017), among others, have highlighted strong concerns with data snooping in the factor-investing literature. Even the usual practice of conducting out-of-sample analysis by dividing data into training and test samples is not immune from data snooping or overfitting. This is because the researcher is observing the test data, and this pseudo out-of-sample testing is also prone to data snooping compared to true out-of-sample testing (Diebold (2015)). I argue that my empirical strategy is akin to a true out-of-sample test as I use the performance after the launch of an index as the out-of-sample period. This data was not available to any user beforehand, thus mainly preventing any look-ahead bias or leakage of future information into the testing process.

Another advantage of using indices instead of factor-mimicking portfolios is that these indices are more tuned to the realistic investment opportunities that an actual investor could have exploited. Such institutional indices exclude stocks that are too small and illiquid. Other issues like rebalancing and weighting are also suitably handled, keeping in mind the interests of actual investors trying to track the index. In contrast, the anomaly literature uses a generic and somewhat ad hoc filtering process along with equal or market cap weighting of stocks. Ledoit, Wolf, and Zhao (2019) highlight that this standard methodology is inefficient at detecting factor performance. Hsu, Kalesnik, and Surti (2010) also argue that market cap weighting dampens the performance of factor portfolios. Another practical benefit of using indices is that they often include additional constraints on a portfolio's individual and/or sectoral concentrations. Such restrictions are relevant to real investors but are usually missing in factor investing studies. Overall, using factor-based indices allows us to use portfolios that are likely to be representative of the returns generated by an investor trying to utilise some factor investing strategy.

Index providers themselves show the performance of their indices in the factsheets and in-house research documents. Then what is the need to conduct a separate analysis of the same? While it is true that any index provider, including Nifty, provides the performance of its indices in its research papers or index related documents, the performance shown is generally for the entire period (i.e., from the base date till the date of analysis). This reporting of the performance for the whole period masks the out-of-sample performance and doesn't identify the differences between the training and test periods[2]. For an illustration, see Figure 1. Panel A of the figure shows the performance of an index, i.e., the Nifty Low Volatility 30, compared to the benchmark, i.e., the Nifty 500. The period is from April 2005 to September 2022. The index was launched in June 2016. Looking at the full performance in Panel A, one may conclude that the index has outperformed the benchmark by a substantial margin. This outperformance seems to continue in the post-launch period (i.e., after June 2016). However, when I divide the total period into the in-sample (from base date to launch date) and out-of-sample (launch date to current date) periods and measure the cumulative performance of both these indices, the previous inference doesn't appear to hold. The full period outperformance seems to be mainly due to the compounding effect of the in-sample outperformance. In the out-of-sample period, the benchmark has beaten the strategy index. Post-launch, the overall performance of both indices is similar, and also their movements appear to be much more correlated.

This case highlights the need to separate the full period of analysis into in-sample and out-of-sample periods before making any judgements on the performance of factor indices (or any other index, for that matter). In this study, I test the performance of nine strategy-based indices of the National Stock Exchange by decomposing their overall performance into training (i.e., in-sample) and test (out-of-sample) periods.

To the best of my knowledge, very few studies use factor indices and divide them into back-test and out-of-sample periods to compare their performance. Even index factsheets[3], while acknowledging that a part of the performance shown is a back-test, do not explicitly show the back-test performance vis-à-vis the out-of-sample performance. Blitz (2016) and Hsu, Kalesnik, and Surti (2010) are noteworthy studies that use factor indices to understand factor investing performance. Both these studies use MSCI Barra and Russell factor indices and test the ability of these indices to outperform. However, they use the data for the entire period, including the back-test period. As I show in this study, using the entire data period can easily mask the underperformance in the out-of-sample period and make it look like the index has also outperformed in the test period. The primary learning is that it is necessary to

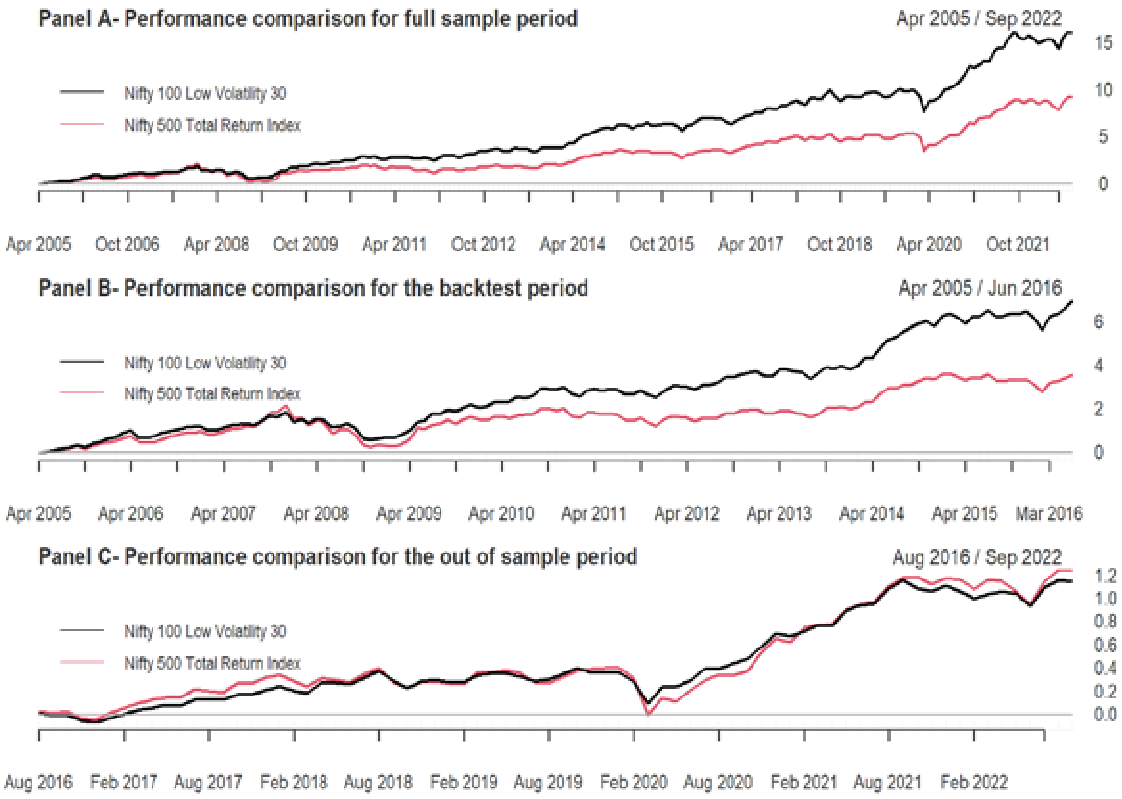
---

[2] I use the terms training period and in-sample interchangeably. Same for test period and out-of-sample.

[3] See https://www.msci.com/documents/10199/4d26c754-8cb9-4fa8-84e6-a51930901367 for an example.

compare the pre-launch performance of an index with its post-launch performance to get a complete and unbiased picture of the performance of a factor strategy.

Figure 1: This figure shows the cumulative performance of the Nifty 100 Low Volatility 30 index compared to the Benchmark Nifty 500 Index.



Note: Panel A shows the performance for the entire sample period, whereas Panels B and C show the performance for the in-sample and out-of-sample periods.

Two closely related studies - Gorman and Fabozzi (2022) and Suhonen, Lennkh, and Perez (2017) use tradable indices to evaluate the out-of-sample performance of factor investing strategies. However, Gorman and Fabozzi (2022) are focused on the period of 2018-2020, during which factor investing was generally going through a rough period. In contrast, my results are for the entire sample period available and not just restricted to the two mentioned years. Suhonen, Lennkh, and Perez (2017), on the other hand, are not restricted to just two years. However, their tests are based on proprietary data of 215 strategies, for which information about the separation of back-testing and live periods was available. Their strategies have a minimum out-of-sample test duration of just .44 years, whereas my study has at least 5 years of out-of-sample data for any strategy. Further, my study uses publicly available data and there is full disclosure about the indices used. Thus, the results in this study are replicable in the spirit of Welch (2019). Further, both these studies use strategies from developed markets. Among other things, this study also tries to understand whether the performance declines reported in the two mentioned studies are also observed in an emerging market.

My study fills the gap in the extant literature by comparing the performance of nine factor-based indices of stocks listed in the Indian stock market. I find that the performance of most indices drop sharply in the out-of-sample period. The results hold for absolute as well as excess and risk-adjusted returns of the indices. Multi-factor analysis of indices suggests that exposure to common factors has increased in the out-of-sample period. The findings are robust to the removal of the COVID-19 period. Overall, these results cast doubt on factor-based indices' ability to generate additional premia consistently.

2. Data and Methodology

The main data source for this study is the National Stock Exchange's (NSE) Nifty indices website. This website provides a list of all nifty indices and the values of the indices from the base date of the index. While there are many strategy-based indices launched by the NSE, my final sample consists of nine of these. I only select indices launched for at least five years as of September 2022. Therefore, only those indices with five years or more of out-of-sample performance data are chosen for the analysis.

Further, I focus on equity indices that follow popular factor investing strategies, such as value, low volatility, momentum, quality etc. This rules out indices of IPOs, futures contracts, and debt securities. The details of the nine chosen indices are in table 1. The data for the launch dates, base dates and rebalancing frequencies have been collected from the respective factsheets of the indices. There is another major provider of strategy-based indices in India, i.e., the ASIA Index Pvt Ltd., which is a joint venture between the Bombay stock exchange (BSE) and S&P Dow Jones indices. However, BSE only provides the data for its indices from the launch date onwards. Because of the lack of data from the base date to the launch date, I do not consider BSE indices in the sample. This exclusion shouldn't affect my inferences as all the major factor strategies are well covered in the NSE indices.

For the risk-free rates, I've collected the data of the Government of India's 10-year bond's monthly yields from the Reserve Bank of India's website. The Fama and French (1993) and Momentum factor data for India are from Agarwalla, Jacob, and Varma (2014). All indices are total return indices, thus including returns adjusted for dividends, stock splits, bonus issues, and similar corporate actions.

Table 1: Names and details of all the nine strategy indices in the sample

Name	Date of Launch	Base Date	Underlying Strategy/Factor	Rebalancing Frequency
Nifty 100 Low Volatility 30	08-07-2016	01-04-2005	Low historical volatility	Quarterly
Nifty Alpha 50	19-11-2012	31-12-2003	Historical alpha	Quarterly
Nifty Dividend Opportunities 50	22-03-2011	01-10-2007	Dividend yield	Annual
Nifty Growth Sectors 15	22-05-2014	01-01-2009	Sectoral P/E and P/B and EPS	Semi-Annual
Nifty High Beta 50	19-11-2012	31-12-2003	Historical beta	Quarterly
Nifty Low Volatility 50	19-11-2012	31-12-2003	Low historical volatility	Quarterly
Nifty 100 Quality 30	19-03-2015	01-10-2009	ROE, Leverage, and EPS Growth	Semi-Annual
Nifty Alpha Quality Low Volatility 30	10-07-2017	01-04-2005	Historical alpha, volatility, and quality scores	Semi-Annual
Nifty Quality Low Volatility 30	10-07-2017	01-04-2005	Historical alpha, volatility, and quality scores	Semi-Annual

Note: This table contains the details of all the nine indices considered in the study. The date of launch is the date when the index was launched for use. The base Date is the first date for which the value of the index is available.

While five years of out-of-sample testing may seem very short, it should be noted that most factor investing indices take factors from academic literature. The literature itself has developed over the last

30 odd years (Since studies like Fama and French (1993) and Jegadeesh and Titman (1993)). Even a tiny out-of-sample period is enough information compared to only the backtest performance. Recently, HDFC mutual fund in India launched an ETF based on the "Nifty 200 Momentum 30" index. The index was launched on August 25, 2020. ETF providers are unlikely to wait long before launching their products. Even 5 years of data seem good given the market context for investors looking to invest in ETFs of such indices. A small out-of-sample period also has some unexpected benefits. There is a higher chance that index providers will tune their methodology over extended periods, making it hard to compare them over time. This is less likely to happen in a shorter duration.

3. Results

In this section, I discuss the findings of the study. For any given index, the full period refers to the duration from the base date of the index up to the end period of data collection, i.e., August 2022. The training period (or the in-sample period) is the period from the base date to the launch date of an index. The test period is the period after the launch up to August 2022. The performance of an index in this period reflects its out-of-sample performance. All the analyses done in this study have been reported for the full training and test periods. Table 2 contains the descriptive statistics of all nine strategy indices for the three periods.

Table 2: Descriptive statistics of the monthly returns of all the nine strategy indices in the sample

	Nifty 100 Low Volatility 30			Nifty Alpha 50			Nifty Dividend Opportunities 50		
	Full	Train	Test	Full	Train	Test	Full	Train	Test
Annualized Return	0.18	0.20	0.13	0.21	0.19	0.22	0.11	0.12	0.11
Observations	209	135	74	224	106	118	179	41	138
Minimum	-0.22	-0.22	-0.15	-0.37	-0.37	-0.24	-0.26	-0.26	-0.15
Maximum	0.18	0.18	0.13	0.29	0.29	0.17	0.30	0.30	0.15
Stdev	0.05	0.06	0.04	0.09	0.10	0.06	0.06	0.10	0.05
Skewness	-0.62	-0.69	-0.43	-0.94	-0.93	-0.68	-0.19	-0.32	0.07
Kurtosis	2.35	1.93	2.16	2.87	1.85	2.06	4.49	1.75	0.68
	Nifty Growth Sectors 15			**Nifty High Beta 50**			**Nifty Low Volatility 50**		
	Full	Train	Test	Full	Train	Test	Full	Train	Test
Annualized Return	0.19	0.31	0.12	0.05	0.06	0.04	0.18	0.20	0.16
Observations	164	64	100	224	106	118	224	106	118
Minimum	-0.23	-0.08	-0.23	-0.37	-0.36	-0.37	-0.22	-0.22	-0.16
Maximum	0.19	0.19	0.17	0.65	0.65	0.31	0.24	0.24	0.13
Stdev	0.05	0.04	0.05	0.11	0.13	0.10	0.05	0.07	0.04
Skewness	-0.43	0.71	-0.86	0.60	0.86	-0.12	-0.57	-0.59	-0.52
Kurtosis	4.63	1.79	5.23	4.47	4.46	1.75	3.08	2.08	1.81
	Nifty 100 Quality 30			**Nifty Alpha Quality Low Vol 30**			**Nifty Quality Low Vol 30**		
	Full	Train	Test	Full	Train	Test	Full	Train	Test
Annualized Return	0.13	0.18	0.09	0.18	0.20	0.12	0.17	0.19	0.12
Observations	153	65	88	209	147	62	209	147	62
Minimum	-0.18	-0.08	-0.18	-0.23	-0.23	-0.15	-0.23	-0.23	-0.13
Maximum	0.12	0.11	0.12	0.14	0.14	0.11	0.16	0.16	0.10
Stdev	0.04	0.04	0.04	0.05	0.05	0.04	0.05	0.05	0.04
Skewness	-0.46	-0.12	-0.67	-1.02	-1.17	-0.47	-0.88	-1.03	-0.26
Kurtosis	1.61	-0.42	2.79	3.37	3.61	1.57	3.35	3.61	1.09

Note: This table contains the basic descriptive statistics for all the indices for three periods. Full refers to the full period for which the index data is available. Train refers to the period from the base date to the launch date of the index. The test period refers to the data after the launch of the index. Returns are in decimals; therefore, .15 means 15%.

From table 2, we can see that seven out of nine indices have shown a marked decline in annualised returns. Only one index, i.e., the Alpha 50, has shown an increase in annualised returns, whereas the dividend opportunities index has shown a very modest decline (less than 100 basis points) in the out-of-sample period. However, the standard deviation of most index returns has also fallen in the test period. Hence, one needs to be careful in making a judgement based on return only as the risk has also fallen. As an additional test, I report the t-tests for the difference in average returns during the training and test periods.

The results in table 3 show that while the returns have fallen for all indices in the test period, none of them are significant at the conventional significance levels. Given that the standard deviation has also fallen during the test period, these results are not entirely surprising. However, these results should not be considered as evidence that there is no significant drop (or increase) in the performance of the aforementioned indices. First, the fall in the standard deviation of returns could be due to a fall in the standard deviation of the underlying factors that drive returns (such as the market factor in the CAPM). Therefore, controlling for the changes in these underlying factors would yield clearer insights into the performance of these indices.

Also, owing to the different time periods for the indices, absolute returns are not directly comparable. Therefore, returns need to be compared to some benchmark and adjusted for risk for a proper comparison between the training and test periods and among each other.

Table 3: Statistical tests of mean difference in average monthly returns

Index	Mean Difference	t-stat
Nifty 100 Low Volatility 30	-0.006	-0.701
Nifty Alpha 50	-0.002	-0.164
Nifty Dividend Opportunities 50	-0.005	-0.429
Nifty Growth Sectors 15	-0.014	-1.678
Nifty High Beta 50	-0.006	-0.366
Nifty Low Volatility 50	-0.004	-0.557
Nifty 100 Quality 30	-0.007	-0.870
Nifty Alpha Quality Low Volatility 30	-0.007	-0.750
Nifty Quality Low Volatility 30	-0.007	-0.764

Note: This table reports the difference between the average returns of the training and test periods. Also reported is the t-statistic of the test under the null hypothesis that the means are the same. All standard errors use the Newey West correction with 4 lags.

Therefore, I estimate adjusted and risk-adjusted returns for the indices along with other indicators of fund performance. These measures are reported in table 4.

I have considered the broad-based NIFTY 500 index as a common benchmark for all our indices. Unreported results are similar with the more popular NIFTY 50 as a benchmark. Table 4 reports the Jensen's alpha, beta, upside and downside betas, Sharpe, modified Sharpe and Treynor ratios of all the indices. The details of the calculation of these indicators are given in table A1 in the appendix.

Even on the basis of risk-adjusted returns, it appears that all indices except the Alpha 50 have shown a decline in the out-of-sample performance. The alpha of the three indices has become negative, suggesting that these indices have underperformed the benchmark on a risk-adjusted basis. The Sharpe and modified Sharpe ratios (using expected shortfall as a risk measure) also tell the same story. All except the alpha 50 index have shown a fall in performance compared to the training period.

While Quality Investing has recently gained some popularity in academic literature, my results show that all three indices with elements of quality investing have underperformed their benchmarks. Further, combining alpha with quality and/or low volatility has also diminished the performance of the alpha strategy, i.e., the only strategy that has worked out-of-sample.

The results until now show that, barring the Alpha 50, all indices have shown a decline in absolute as well as relative performance compared to a benchmark.

A typical factsheet provided by the index provider starts measuring the index's performance from the base date. Regular updates to these factsheets keep on adding performance data as the timeline progresses. However, the full period performance still contains the training period performance. A key takeaway from the results is that by looking at the full period performance, an investor is likely to overestimate the expected future returns from a strategy. It will be useful for investors if index providers separate the back-test performance from the actual out-of-sample performance of a factor index.

Table 4: Indicators of the indices' adjusted returns, risk, and risk-adjusted returns.

	Nifty 100 Low Volatility 30			Nifty Alpha 50			Nifty Dividend Opportunities 50		
	Full	Train	Test	Full	Train	Test	Full	Train	Test
Active Return	0.034	0.059	-0.009	0.066	0.049	0.081	0.017	0.106	-0.011
Annualized Alpha	0.043	0.064	0.009	0.063	0.054	0.074	0.017	0.097	-0.004
Beta	0.747	0.756	0.714	1.149	1.169	1.101	0.874	0.898	0.846
Beta+	0.663	0.633	0.755	0.923	0.881	0.957	0.845	0.806	0.863
Beta-	0.794	0.870	0.633	1.350	1.408	1.226	0.848	1.094	0.668
R-squared	0.870	0.876	0.857	0.823	0.866	0.725	0.890	0.929	0.847
Treynor Ratio	0.127	0.152	0.083	0.110	0.095	0.127	0.041	0.046	0.041
StdDev Sharpe	0.169	0.183	0.138	0.162	0.140	0.203	0.079	0.084	0.084
ES Sharpe (99%)	0.055	0.055	0.045	0.051	0.036	0.063	0.024	0.025	0.030
	Nifty Growth Sectors 15			**Nifty High Beta 50**			**Nifty Low Volatility 50**		
	Full	**Train**	**Test**	**Full**	**Train**	**Test**	**Full**	**Train**	**Test**
Active Return	0.035	0.129	-0.021	-0.092	-0.086	-0.097	0.033	0.055	0.014
Annualized Alpha	0.063	0.177	-0.008	-0.084	-0.072	-0.100	0.040	0.058	0.026
Beta	0.595	0.369	0.864	1.583	1.520	1.740	0.779	0.784	0.766
Beta+	0.387	0.226	0.825	1.841	1.830	1.987	0.708	0.720	0.652
Beta-	0.803	0.658	0.905	1.307	1.152	1.633	0.818	0.888	0.736
R-squared	0.522	0.338	0.754	0.859	0.905	0.793	0.917	0.935	0.874
Treynor Ratio	0.179	0.582	0.049	-0.014	-0.010	-0.016	0.123	0.149	0.100
StdDev Sharpe	0.201	0.394	0.095	0.040	0.054	0.025	0.167	0.172	0.173
ES Sharpe (99%)	0.061	0.268	0.026	0.007	0.010	0.007	0.055	0.054	0.054
	Nifty 100 Quality 30			**Nifty Alpha Quality Low Vol 30**			**Nifty Quality Low Vol 30**		
	Full	**Train**	**Test**	**Full**	**Train**	**Test**	**Full**	**Train**	**Test**
Active Return	0.009	0.058	-0.026	0.035	0.054	-0.008	0.029	0.045	-0.008
Annualized Alpha	0.018	0.064	-0.015	0.048	0.066	0.008	0.044	0.059	0.009
Beta	0.742	0.693	0.776	0.684	0.685	0.682	0.663	0.667	0.649
Beta+	0.661	0.517	0.775	0.514	0.475	0.696	0.535	0.510	0.656
Beta-	0.728	0.743	0.730	0.805	0.872	0.657	0.756	0.830	0.589
R-squared	0.772	0.681	0.845	0.808	0.819	0.773	0.807	0.820	0.766
Treynor Ratio	0.068	0.130	0.028	0.139	0.170	0.069	0.136	0.163	0.072
StdDev Sharpe	0.117	0.191	0.063	0.176	0.199	0.110	0.171	0.191	0.113
ES Sharpe (99%)	0.036	0.085	0.020	0.055	0.062	0.034	0.054	0.060	0.036

Note: This table contains the regular, upside and downside betas of the indices. The r-squared with the benchmark is also shown. Three risk-adjusted performance measures are also given- Treynor, Sharpe, and modified Sharpe ratios. The modified Sharpe ratio uses the 99% expected shortfall (also known as conditional Value at Risk) as a risk measure instead of the standard deviation.

For the next set of analyses, I use multi-factor regressions of the following form to check whether any of the indices generate significant abnormal returns after controlling for exposures to the market, size, value, and momentum factors.

$$R_t - Rf_t = \alpha + \beta_{MF} * MF_t + \beta_{SMB} * SMB_t + \beta_{HML} * HML_t + \beta_{WML} * WML_t + \varepsilon_t \quad (1)$$

Where $R_t - Rf_t$ refers to the excess return on an index at the time 't'. MF_t, SMB_t, HML_t, WML_t refer to the returns on the market, size, value, and momentum factors at time 't'.

The results for all indices for all three periods have been reported in table 5. Taking t-stat >2 as a benchmark for statistical significance, around 4 out of 9 indices generated significant alpha in the training period. However, none of the indices generated significant abnormal returns in the test period. Even the Alpha 50 index's abnormal returns are insignificant after controlling for multiple factors. Seven indices have an increased loading on the market factor in the test period compared to the training period. The average increase in the market beta for all the indices is around .09. Therefore, indices are more exposed to market movements in the testing period than when the back-test was done. The proclaimed benefits of providing countercyclical exposures don't seem to have materialised. One of the major factors behind the decline of the performance of the Nifty growth sectors 15 index is the increase in exposure to market risk. The three indices based on quality investing had modest returns in the training sample, but their performance never really took off in the test period.

Table 5: Results of the multi-factor regressions of Index returns.

	Nifty 100 Low Volatility 30					
	Full		**Train**		**Test**	
	Coef.	**t-stat**	**Coef.**	**t-stat**	**Coef.**	**t-stat**
Alpha	0.003	2.278	0.004	2.944	-0.001	-0.786
SMB	-0.088	-1.806	-0.094	-1.502	-0.151	-3.971
HML	-0.027	-0.766	-0.007	-0.163	-0.178	-5.917
WML	0.024	0.428	0.034	0.531	-0.063	-1.620
MF	0.771	13.045	0.758	11.600	0.917	28.767
R-squared	0.863		0.858		0.930	
	Nifty Alpha 50					
	Full		**Train**		**Test**	
	Coef.	**t-stat**	**Coef.**	**t-stat**	**Coef.**	**t-stat**
Alpha	0.002	0.928	0.001	0.438	0.002	0.965
SMB	0.341	6.619	0.366	5.380	0.332	4.277
HML	-0.013	-0.222	-0.086	-1.084	0.068	1.118
WML	0.192	3.099	0.191	2.267	0.225	3.054
MF	1.153	17.411	1.147	15.261	1.167	20.631
R-squared	0.883		0.898		0.851	
	Nifty Dividend Opportunities 50					
	Full		**Train**		**Test**	
	Coef.	**t-stat**	**Coef.**	**t-stat**	**Coef.**	**t-stat**
Alpha	0.002	0.881	0.006	1.587	0.000	-0.184
SMB	-0.020	-0.332	0.172	2.091	-0.126	-2.808
HML	0.123	2.636	0.112	1.005	0.098	2.454
WML	-0.069	-1.419	-0.050	-0.636	-0.047	-0.993
MF	0.806	13.328	0.782	11.371	0.861	19.407
R-squared	0.877		0.916		0.843	

	Nifty Growth Sectors 15					
	Full		Train		Test	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Alpha	0.003	1.036	0.011	2.504	0.002	0.413
SMB	-0.004	-0.046	-0.242	-2.471	0.023	0.223
HML	-0.004	-0.040	-0.194	-2.637	-0.161	-1.562
WML	0.132	1.938	0.180	2.156	-0.226	-2.280
MF	0.684	5.464	0.630	4.686	0.913	10.080
R-squared	0.479		0.474		0.710	
	Nifty High Beta 50					
	Full		Train		Test	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Alpha	-0.002	-0.951	-0.001	-0.145	-0.003	-1.018
SMB	0.087	1.113	0.075	0.839	0.183	1.330
HML	0.253	3.275	0.057	0.682	0.490	5.770
WML	-0.448	-8.572	-0.446	-5.867	-0.389	-4.546
MF	1.334	25.952	1.333	23.445	1.290	11.158
R-squared	0.895		0.934		0.848	
	Nifty Low Volatility 50					
	Full		Train		Test	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Alpha	0.003	2.823	0.004	2.856	0.001	0.690
SMB	0.014	0.380	-0.009	-0.197	0.012	0.302
HML	-0.032	-1.039	-0.032	-0.755	-0.069	-1.294
WML	0.016	0.401	0.006	0.106	0.034	1.041
MF	0.788	19.857	0.765	17.078	0.901	35.404
R-squared	0.911		0.920		0.899	
	Nifty 100 Quality 30					
	Full		Train		Test	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Alpha	0.001	0.719	0.004	2.395	-0.001	-0.334
SMB	-0.104	-1.691	-0.258	-3.849	-0.053	-0.763
HML	-0.168	-3.763	-0.187	-3.685	-0.233	-5.354
WML	0.040	0.739	0.145	2.611	-0.140	-2.591
MF	0.939	24.421	0.999	18.199	0.922	25.349
R-squared	0.809		0.812		0.882	
	Nifty Alpha Quality Low Vol 30					
	Full		Train		Test	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Alpha	0.002	1.896	0.004	2.464	-0.002	-0.818
SMB	0.024	0.516	-0.002	-0.027	0.006	0.143
HML	-0.077	-2.108	-0.069	-1.642	-0.266	-4.859
WML	0.125	2.779	0.140	2.725	-0.030	-0.590
MF	0.739	15.072	0.728	13.147	0.875	35.236
R-squared	0.853		0.853		0.911	
	Nifty Quality Low Vol 30					
	Full		Train		Test	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Alpha	0.003	2.448	0.004	2.862	-0.001	-0.625
SMB	-0.012	-0.200	-0.027	-0.374	-0.066	-1.176
HML	-0.065	-1.480	-0.044	-0.900	-0.291	-5.088
WML	0.025	0.439	0.036	0.548	-0.122	-2.580
MF	0.683	11.158	0.673	9.825	0.827	23.094
R-squared	0.815		0.812		0.891	

Note: This table reports the coefficients and t-statistics of the intercept and the loadings of each index on the market, size, value, and momentum factors. All t-statistics are based on the Newey west correction with 4 lags.

Table 6: Results of the multi-factor regressions of Index returns with dummy variable for the test period.

	Incl. COVID Period			Excl. COVID Period		
	Coef.	t-stat	p-val	Coef.	t-stat	p-val
Nifty 100 Low Volatility 30						
Alpha	0.005	2.507	0.013	0.005	3.132	0.002
OOS_dummy	-0.006	-1.801	0.073	-0.006	-2.027	0.044
Nifty Alpha 50						
Alpha	0.001	0.193	0.847	0.001	0.222	0.824
OOS_dummy	0.003	0.610	0.543	0.003	0.826	0.410
Nifty Dividend Opportunities 50						
Alpha	0.006	1.591	0.114	0.006	1.332	0.185
OOS_dummy	-0.006	-1.316	0.190	-0.006	-1.106	0.270
Nifty Growth Sectors 15						
Alpha	0.010	2.200	0.030	0.010	3.090	0.002
OOS_dummy	-0.013	-2.114	0.036	-0.013	-2.528	0.013
Nifty High Beta 50						
Alpha	-0.002	-0.613	0.541	-0.002	-0.675	0.501
OOS_dummy	0.000	0.023	0.982	-0.001	-0.120	0.905
Nifty Low Volatility 50						
Alpha	0.004	2.386	0.018	0.004	2.696	0.008
OOS_dummy	-0.002	-0.933	0.352	-0.002	-0.961	0.338
Nifty 100 Quality 30						
Alpha	0.006	2.367	0.019	0.006	2.613	0.010
OOS_dummy	-0.009	-2.782	0.006	-0.009	-3.039	0.003
Nifty Alpha Quality Low Vol 30						
Alpha	0.004	2.152	0.033	0.004	2.630	0.009
OOS_dummy	-0.006	-1.707	0.090	-0.006	-1.841	0.067
Nifty Quality Low Vol 30						
Alpha	0.005	2.454	0.015	0.005	3.083	0.002
OOS_dummy	-0.006	-1.578	0.116	-0.005	-1.513	0.132

Note: This table reports the coefficients and t-statistics of the intercept and a dummy variable for the test period. All t-statistics are based on the Newey west correction with 4 lags.

Multi-factor regressions show that while five indices generate significant alpha at the 10% level in the training period, not a single index has significant alpha in the test period. These results hint at a substantial decline in the performance of indices after their launch. To test whether the declines in the alpha are statistically significant, I rerun the regression above using the entire sample and using an additional dummy variable which takes the value of 1 for the test period and zero otherwise[4]. These results are reported in table 6. As an additional robustness test, I also run the same analysis after removing the COVID-19 period.[5] For brevity, only the results of the focus variable (i.e., the test period dummy) have been reported.

A negative and significant value for the dummy coefficient shows that alpha has fallen significantly in the out-of-sample period. Of the nine indices, four show a significant decline in the alphas in the test period. None of the indices shows a significant increase in alpha. Excluding the COVID-19 period doesn't change the inferences. If anything, the significance is higher in the filtered sample.

Further, in the four indices where there is a significant decline in the alpha, the average R-squared with the four factors has increased by around 10%. Therefore, the correlation with existing factors increased during the test period. Linnainmaa and Roberts (2018) show that this higher correlation in the out-of-sample period is likely to be an artefact of data snooping. Also, most indices show a higher exposure to the market in the test period compared to the training period. While our tests do not have the power to differentiate between multiple explanations of this phenomenon, the findings are nonetheless consistent with the explanation that the training sample's abnormal returns might be a result of excessive data snooping.

One possible alternative explanation of our results is that due to the small out-of-sample period, some of our tests may have limited power to detect abnormal performance in the test period, even if it existed. As per the results in table 6, it seems unlikely because 4 out of nine indices show a significant decline in alphas. If anything, the lower power of the tests would likely work in favour of the indices, with us being unable to reject the null of no underperformance. Nonetheless, to further assuage concerns regarding a small test sample, I use a placebo test[6] in which I consider a five-year window from the training sample of each of the indices. The placebo window consists of the last five years of data from the original training sample. Using this placebo data, I rerun the multi-factor tests reported in table 5. The alpha coefficients of the multi-factor regressions are reported in table 7.

Table 7: Results of the multi-factor regressions of Index returns using a placebo period.

Index	Alpha coefficient	t-stat	p-val
Nifty 100 Low Volatility 30	0.003	1.856	0.069
Nifty Alpha 50	-0.001	-0.232	0.818
Nifty Dividend Opportunities 50	0.006	1.587	0.121
Nifty Growth Sectors 15	0.012	2.528	0.014
Nifty High Beta 50	0.000	0.071	0.944
Nifty Low Volatility 50	0.006	3.137	0.003
Nifty 100 Quality 30	0.005	2.682	0.010
Nifty Alpha Quality Low Volatility 30	0.002	1.236	0.222
Nifty Quality Low Volatility 30	0.001	0.814	0.419

Note: This table reports the coefficients and t-statistics of the intercept of the multi-factor regressions using a placebo test period. All t-statistics are based on the Newey west correction with 4 lags.

These results show that even using just a five-year period from the training data, four out of nine indices show a significant alpha. Thus, a five-year period seems sufficient enough to detect abnormal

[4] I thank two anonymous referees for suggesting this analysis.
[5] I define the COVID-19 period as February to October 2020. This was the main period of the extreme stock events (crash and subsequent recovery) during COVID-19. Major stock indices had recovered to levels closer to January 2020 prices by November 2020.
[6] I thank an anonymous referee for this suggestion.

performance. The power of our tests using smaller samples may not be ideal, but it seems unlikely that the results are entirely due to statistical noise.

To summarise, there are six general anomalies or factor strategies represented in these nine indices – quality, value, low volatility, high beta, and momentum (historical alpha). Based on the analysis, it doesn't seem that any of these anomalies are robust enough in out-of-sample analysis. Studies such as Linnainmaa and Roberts (2018), Mclean and Pontiff (2016), and Hou, Xue, and Zhang (2020) have shown that the performance of factor strategies tends to decline in the periods after they have been observed. Using tradable factor indices, - Gorman and Fabozzi (2022) and Suhonen, Lennkh, and Perez (2017) also report a decline in the out-of-sample performance of factor investing strategies.

Ultimately, my results, combined with the findings of these studies, show that the factor investing hype is yet to live up to its promise. That said, not all is damning for factor investing. Ledoit, Wolf, and Zhao (2019) and Hsu, Kalesnik, and Surti (2010) have shown that factor underperformance may be tackled by using more sophisticated weighting criteria. Amenc et al. (2015) also recognise that index providers do not always efficiently deal with the issue of ensuring out-of-sample robustness of factor investing strategies. They suggest certain practices that index creators can follow to improve the robustness of factor portfolios. While factor investing holds promise, a lot more effort must be put into issues like weighting, model overfitting, and exposure to existing factors to ensure consistency in performance.

4. Conclusion

I test the robustness of factor investing strategies by analysing the returns of factor-based indices after their launch and comparing them with their pre-launch performance. The results show an evident decline in the performance of most strategy indices compared to their back-test performance. Barring one index, i.e., the Alpha 50, all of the indices underperform the benchmark out-of-sample. The results cast considerable doubt on the ability of factor investing to generate excess returns.

These results are beneficial for investors and academicians attracted to factor investing. Despite having other potential benefits, the main selling proposition of factor funds is their outperformance. The awareness that past outperformance has not held up in the future can help investors make better investment decisions.

References

Agarwalla, Sobhesh Kumar, Joshy Jacob, and Jayanth Rama Varma, 2014, Four factor model in Indian equities market, Indian Institute of Management, Ahmedabad Working Paper, 05.

Amenc, Noël, Felix Goltz, Sivagaminathan Sivasubramanian, and Ashish Lodh, 2015, Robustness of Smart Beta Strategies, The Journal of Beta Investment Strategies 6, 17–38.

Banz, Rolf W., 1981, The relationship between return and market value of common stocks, Journal of Financial Economics 9, 3–18.

Blitz, David, 2016, Factor Investing with Smart Beta Indices, The Journal of Beta Investment Strategies 7, 43–48.

Cakici, Nusret, Adam Zaremba, Robert J. Bianchi, and Nga Pham, 2021, False discoveries in the anomaly research: New insights from the Stock Exchange of Melbourne (1927–1987), Pacific-Basin Finance Journal 70, 101675.

Diebold, Francis X., 2015, Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests, Journal of Business & Economic Statistics 33, 1–1.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, Journal of Financial Economics 33, 3–56.

Gorman, Stephen A., and Frank J. Fabozzi, 2022, Workhorse or Trojan Horse? The Alternative Risk Premium Conundrum in Multi-Asset Portfolios, The Journal of Portfolio Management 48, 147–182.

Harvey, Campbell R., 2017, Presidential address: The scientific outlook in financial economics, The Journal of Finance 72, 1399–1440.

Hollstein, Fabian, 2022, The world of anomalies: Smaller than we think? Journal of International Money and Finance 129, 102741.

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, The Review of Financial Studies 33, 2019–2133.

Hsu, Jason, Vitali Kalesnik, and Himanshu Surti, 2010, An Examination of Traditional Style Indices, The Journal of Beta Investment Strategies 1, 14–23.

Huang, Jing-Zhi, and Zhijian Huang, 2014, Real-Time Profitability of Published Anomalies: An Out-of-Sample Test, The Quarterly Journal of Finance.

Jacobs, Heiko, and Sebastian Müller, 2020, Anomalies across the globe: Once public, no longer existent? Journal of Financial Economics 135, 213–230.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, The Journal of finance 48, 65–91.

Ledoit, Olivier, Michael Wolf, and Zhao Zhao, 2019, Efficient Sorting: A More Powerful Test for Cross-Sectional Anomalies, Journal of Financial Econometrics 17, 645–686.

Linnainmaa, Juhani T, and Michael R Roberts, 2018, The History of the Cross-Section of Stock Returns, The Review of Financial Studies 31, 2606–2649.

Lo, Andrew W., and A. Craig MacKinlay, 1990, Data-Snooping Biases in Tests of Financial Asset Pricing Models, The Review of Financial Studies 3, 431–467.

Mclean, R. David, and Jeffrey Pontiff, 2016, Does Academic Research Destroy Stock Return Predictability? The Journal of Finance 71, 5–32.

Suhonen, Antti, Matthias Lennkh, and Fabrice Perez, 2017, Quantifying backtest overfitting in alternative beta strategies, The Journal of Portfolio Management 43, 90–104.

Welch, Ivo, 2019, An Opinionated FAQ, Critical Finance Review 8, 19–24.

Appendix

Table A1: Calculation of risk and performance metrics

Indicator	Method of calculation
Active Return	Return on the index minus the benchmark returns
Beta	Covariance of index and benchmark returns divided by the variance of benchmark returns
Beta+	The beta of the index considering only the subsample in which benchmark returns are positive
Beta-	The beta of the index considering only the subsample in which benchmark returns are negative
R-squared	R-squared of the regression of index excess returns on market excess returns
Treynor Ratio	The average return on the index minus the benchmark return, divided by the beta of the index
StdDev Sharpe	The average return on the index minus the risk-free return, divided by their standard deviation
ES Sharpe (99%)	The average return on the index minus the risk-free return, divided by the 99% expected shortfall calculated using the historical simulation method

Table A2: Additional details about the indices

Name	Weighting Scheme	Long/Short	# of Constituents
Nifty 100 Low Volatility 30	Volatility based weighting	Long	30
Nifty Alpha 50	Alpha based weighting	Long	50
Nifty Dividend Opportunities 50	Periodic Capped Free Float Market Cap	Long	50
Nifty Growth Sectors 15	Periodic Capped Free Float Market Cap	Long	15
Nifty High Beta 50	Beta based weighting	Long	50
Nifty Low Volatility 50	Volatility based weighting	Long	50
Nifty 100 Quality 30	Combination of quality score and free float market capitalisation.	Long	30
Nifty Alpha Quality Low Volatility 30	Multi-factor score weighted	Long	30
Nifty Quality Low Volatility 30	Multi-factor score weighted	Long	30

Note: This table contains additional details about the indices used in the study. Further information about current constituents and other financial metrics can be obtained from the respective index factsheets from the website niftyindices.com