Style Investing with Machine Learning

Philipp Kallerhoff[1,2]

[1]Vontobel Asset Management AG, Gotthardstrasse 27, 8022 Zürich, Switzerland

[2]University of Zürich, Rämistrasse 71, 8006 Zürich, Switzerland

Correspondence: Philipp Kallerhoff, Vontobel Asset Management AG, Gotthardstrasse 27, 8022 Zürich, Switzerland. E-mail: pkaller@gmail.com

Abstract

This paper applies machine learning techniques to style investing. Support Vector Regression is applied to multi-factor investing based on momentum, dividend, quality, volatility and growth. The results show that Support Vector Regression selects stocks consistently with a higher efficiency ratio than a broad market investment and outperforms linear regression methods. The methods are applied to global stocks in the MSCI World index between 1996 and 2016. The behavior of both models is analyzed for economic sectors and over time. Interestingly, factors like low-volatility and momentum contribute both positively and negatively in some economic sectors and certain time periods.

1. Introduction

Style investing has been studied extensively in equity markets at least since Fama and French (1992). For this approach, stocks from a broad index are grouped together with similar characteristics. Fama and French (1992) combine two styles, value and size, to capture the cross-sectional variation in average stock returns. The list of factors has been extended over the past. Jegadeesh and Titman (1993) show that buying stocks that performed well in the past generate significant outperformance over the broad index. Ang et al. (2006) and Li et al. (2016) find that stocks with high idiosyncratic volatility have low average returns.

Although the amount and explicit factor for an optimal solution remains under research, this paper focuses on growth, momentum, quality, volatility and dividend factors. Asness et al. (2015) showed that these variables have success in asset pricing. They have the advantage of a strong economic intuition and track record of producing long-term positive returns. The factors can also be traded long and short to extract the risk premium from the market and to produce uncorrelated returns to long-only asset classes. This paper excludes common factors like size and illiquidity, which show a smaller premium and are less scalable.

Investing in these factors is referred to as style investing (Asness et al., 2015). One way of investing is to sort stocks according to their factor values. This univariate sorting procedure (see e. g. Fama & French, 1993) buys the stocks with the highest factor values and sells the stocks with the lowest factor values. Combining several styles on the same set of stocks would lead to offsetting trades on some stocks and aggregating long or short trades on other stocks.

Another way of investing applies multivariate regression to style investing (e. g. Haugen & Baker, 1996; Hanna & Ready, 2005; Fama & French, 2008; Lewellen, 2014). This approach regularly calculates a regression between factor values and future stock returns. The stocks are sorted with their factor values multiplied with the regression coefficients. Both procedures can be repeated over time and scaled to a target volatility.

This paper applies machine learning to style investing. Machine learning has been very successful in various fields such as search engines and computer vision. We apply machine learning to find an optimal combination between the different styles instead of applying a univariate sorting procedure or simple regression. In particular, we apply Support Vector Machines (SVM) (Cortes & Vapnik, 1995), which are among the most widely used algorithms in machine learning.

We compare our results with linear regression models similar to Novy-Marx (2013) and Fama and French (2015).

These regression models use month-by-month multivariate cross-sectional regressions of key stock characteristics. We apply the Support Vector Machines with the same logic and then compare the performance of each approach.

In this paper we do not focus on factor selection or adding new factors. However, we will provide some insights on how the impact of factors is changing over time and especially in different economic sectors. Therefore, we build the linear and SVM regression for each economic sector separately and repeatedly. We extract factor values for each stock and train both the linear and SVM regression in a single economic sector. As a result we retrieve models for each economic sector that are changing their weights to each factor over time. With this approach, we receive detailed insights, how each of the different factors impacts the performance of stocks over time and in different sectors.

This paper adds to the growing literature of style investing, by applying machine learning models. We compare the behavior of these novel techniques in the setting of well-described linear methods. This paper discusses the combination, but not the timing of factor styles. The regression models weigh the different factors differently over time, but factors are not excluded at any point in time. Of course, training the regression and machine learning models, the factors change their sign in some sectors and at some points in time. These factor rotations can be easily explained, for example, by the technology bubble in 2000, where growth was favored over value in the technology sector. These rotations are learned by the models and reflected in the weighting of factors.

2. Data

We select all stocks from the MSCI World between 07/31/1996 and 06/30/2016 (total of 3506 stocks). For these stocks, we download daily data shown in Table 1 from Bloomberg. All data is downloaded in USD.[1] The total amount of data is 150 million data points including the weight of each stock in the MSCI World index. After downloading, the data is reduced to end of week data for further analysis[2] and the fundamental data (dividend, growth and quality factors) is lagged over 6 months to ensure the data is available at the time of trading.

Table 1. Name and Description of input factors used for modelling future returns

Name	Bloomberg field	Bloomberg field definition	Short name
Price	cumulative_tot_return_gross_dvds	"One day total return as of today. The start date is one day prior to the end date (as of date). Gross dividends are used."	P
Dividend	eqy_dvd_yld_12m	"Sum of gross dividend per share amounts that have gone ex-dividend over the prior 12 months, divided by the current stock price. Gross and Net Dividend amounts are assumed to be the same when only one is reported. All Cash Dividend Types are included in this yield calculation."	DVD
Momentum	cum_tot_ret_6m	6-month total return calculated from the price above.[3]	MOM
Growth	px_to_book_ratio	"Ratio of the stock price to the book value per share. Calculated as: Price to Book Ratio = Last Price / Book Value Per Share."	PB
Quality	return_com_eqy	"Measure of a corporation's profitability by revealing how much profit a company generates with the money shareholders have invested, in percentage. Calculated as: (T12 Net Income Available for Common Shareholders / Average Total Common Equity) * 100."	ROE
Volatility	volatility_260d	"A measure of the risk of price moves for a security calculated from the standard deviation of day to day logarithmic historical price changes. The 260-day price volatility equals the annualized standard deviation of the relative price change for the 260 most recent trading days closing price, expressed as a percentage. When looking at current value, the last price point will be the most recently traded price of the security."	VOL

After downloading the data, the data is separated for each sector and vectorized. The machine learning methods

[1]The currency setting impacts the price, momentum and volatility data, but has no impact on dividends, growth and quality values.

[2]The last day of the week was chosen as Friday. If the data is not available on Friday, then the data was chosen from the last available day of the week unless no data was available for the whole week.

[3]Bloomberg does not supply this field and we calculated this value from the price field

below are agnostic to the time dimension and consider every data point separately. However, each data point can integrate information about past performance such as the momentum indicator above.

The data is analyzed separately for each economic sector. We use the Bloomberg Industry Classification System level I to attribute stocks to economic sectors.

To train the models, each data point was labelled with the one month future return of the stock, because we want to rebalance the portfolio on a monthly basis and select the best performing stocks for this period.

3. Models

First, we utilize a linear prediction model and perform a multiple linear regression to model the relationship between input factors (see table above) and predicted returns r of the corresponding stock i at time t+1 month:

$$r_{i,t+1} = \alpha + \beta_{1,t}DVD_{i,t} + \beta_{2,t}MOM_{i,t} + \beta_{3,t}PB_{i,t} + \beta_{4,t}ROE_{i,t} + \beta_{5,t}VOL_{i,t}$$

The linear model considers an intercept α and linear terms β for each predictor, which in our case comprise of the five input factors discussed above.

Secondly, we train a support vector regression (SVR) model. SVR models are part of kernel techniques that transform the feature space with a kernel and then perform a maximum margin classification. With the Gaussian kernel G the similarity between data points X and the support vectors decreases exponentially. The steepness is controlled by the parameter σ, which we kept at 1.

$$G(X,\sigma) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^5} e^{\frac{-|X|^2}{2\sigma^2}}, X = \begin{pmatrix} DVD_{i,t} \\ MOM_{i,t} \\ PB_{i,t} \\ ROE_{i,t} \\ VOL_{i,t} \end{pmatrix}$$

The goal of the method is to find a function that deviates no more than epsilon for each training point and is at the same time as flat as possible. Two parameters, epsilon and c control this trade-off, which are chosen depending on the variance of the output variable. To penalize misclassifications, SVR impose a cost c for observations outside the epsilon margin. The box constraint was kept at the interquartile range of the future performance in the training set.

The epsilon parameter helps toward a flat function by noise reduction as SVR methods ignore errors in the bandwidth of epsilon. We kept epsilon at the interquartile range of the future performance in the training set.

4. Results (Quintile)

In our first test, the model is trained on 80% of the data and tested on the remaining 20% of the data for each cross-validation. The data is randomly selected over time points and stocks for the training or test data set.

The results are divided into two periods. One period comprises the data between 07/31/1996 and 07/31/2006, and the other the data between 07/31/2006 and 06/30/2016.

The trained models predict the return over the next month for each data point in the test set. These predicted returns are sorted for each sector into quintiles from the lowest predicted returns $\hat{\mu}_{1,min}$ to the highest predicted returns $\hat{\mu}_{10,max}$.

Our null hypothesis is that the actual average return of the lowest quintile $\mu_{1,min}$ is higher than in the highest quintile $\mu_{5,max}$:

$$H0: \mu_{5,max} \leq \mu_{1,min}$$

To test our hypothesis, we use a two-sample t-test to test the alternative hypothesis that the population mean of $\mu_{1,min}$ is less than the population mean of $\mu_{5,max}$:

$$H1: \mu_{5,max} > \mu_{1,min}$$

For this analysis, the significance level is set to 5% and the test performed for each sector, two time periods and 1000 random selected cross-validations for bootstrapping (Efron, 1979). The advantage of this approach is repeated testing of our methods in order to estimate confidence intervals of our results.

We find that the SVR always significantly separates the highest and lowest quintile correctly across sectors, time and cross-validations, whereas the linear models fails in some sectors and especially in the more recent time frame (see Appendix A).

Table 2 shows the average efficiency rate of each quintile averaged over sectors and 1000 cross-validations. For both models, the average performance is monotonously increasing in both time periods. However, the SVR model improves the efficiency of the linear model dramatically. The linear model achieves an efficiency of 0.79 in the highest quintile and the earlier time period. This resembles other results nicely such as Fama and French (2008). Interestingly, the SVR improves the efficiency to 1.28 in the highest quintile and same time period.

Additionally, the SVR increases the difference in efficiency between the highest and lowest quintile. The difference is interesting for the implementation of long and short trades. With the linear model, the difference in efficiency between the highest and lowest quintile is 0.63 in the later time period. The SVR model more than doubles the difference in efficiency to 1.71.

Table 2. Average efficiency rates of the linear versus the SVR model for different quintiles and time periods

		Quintile				
		1	2	3	4	5
Linear model	07/31/1996 to 07/31/2006	0.07	0.18	0.35	0.57	0.79
	07/31/2006 to 06/30/2016	-0.10	0.11	0.26	0.44	0.53
SVR model	07/31/1996 to 07/31/2006	-0.46	0.09	0.33	0.65	1.28
	07/31/2006 to 06/30/2016	-0.57	-0.12	0.25	0.56	1.14

Table 3 shows the average t-statistics in both periods for the linear model. The t-statistics are calculated for a test that the coefficient is zero and are averaged for each sector over 1000 cross-validations. In general, the factors contribute to the sector models as expected from the literature (e.g. Asness et al., 2015). High Dividend yields, high momentum and high quality contribute positively to the future performance of stocks. On the other hand, high growth (low value) and high volatility contribute negatively.

To test the significance of the t-statistics, we calculate confidence intervals [0.025, 0.975] for each sector and period over 1000 cross-validations as suggested by Hall (1988). Significant values are indicated with * in Table 3. Interestingly, the t-statistics show strong significance for the volatility indicator in both periods and across sectors as expected from other research such as Li (2016).

The dividend, growth and quality indicators remain at similar average values of 2.26, -1.33 and 1.26 across the periods. Only the momentum indicator shifts from a slightly negative average t-statistic to a positive t-statistic. This indicates that in the period between 1996 and 2006, mean reversion has dominated the next month return in some sectors, but this changed dramatically to momentum in the period between 2006 and 2016.

The factor contribution varies widely between the different sectors. For example, basic materials stocks are driven by high dividends and high momentum, whereas financial stocks are driven by low volatility.

Table 3. Average t-statistics of the linear regression model for two time periods

	07/31/1996 to 07/31/2006					07/31/2006 to 06/30/2016				
	Dividend	Momentum	Growth	Quality	Volatility	Dividend	Momentum	Growth	Quality	Volatility
Basic Materials	4.88*	2.63*	-2.92*	0.67	-3.36*	1.76	5.36*	-1.38	1.78	-4.57*
Communications	7.15*	-4.42*	-1.75*	2.92*	-7.91*	-0.44	4.84*	-2.42*	-0.01	-2.90*
Consumer, Cyclical	5.00*	-1.46*	-0.38	0.53	-7.89*	3.21*	9.02*	-0.14	-0.31	-1.95
Consumer, Non-cyclical	4.18*	0.89	-4.60*	5.43*	-8.49*	2.77	6.40*	0.90	0.79	-5.85*
Diversified	3.29*	-0.63	-1.49*	2.61*	-2.76*	-1.11	1.64	-2.58*	1.95	-0.03
Energy	2.92*	-0.44	0.52	2.91*	-8.09*	1.64	4.89*	3.21*	-1.27	-2.16
Financial	0.67	-1.82*	-0.37	-1.25*	-15.83*	1.05	8.50*	-2.92*	0.73	-3.93*
Industrials	1.33	-0.04	-2.78*	0.15	-12.16*	0.16	7.51*	-0.50	2.58*	-3.46*
Technology	0.68	1.30	-3.05*	2.88*	-1.87*	0.56	3.21*	-1.60	1.87	-1.44
Utilities	3.21*	1.48	-0.67	-0.13	-7.8*	2.25*	4.67*	-1.57	0.34	-3.18*

The SVR model does not depend on single coefficients such as the linear model, but rather on a function of the input variables. This function is defined by the Gaussian Kernel, which limits the non-linearity of the response. In comparison to the linear model, the SVR model disregards extreme factor values and does not respond to them. For example, the most recent SVR model (trained on 01/01/2016) does not respond to price-to-book values above five in the utility sector. For other values, the SVR model responds similar to the linear model. In particular, the SVR model shows clear positive dependency on dividends, momentum and quality as well as negative dependency on volatility and growth (see Appendix B).

5. Results (Rolling)

For the second test, we train the linear and SVR models over a rolling period of time and then use the model to build the portfolio in the next period. Again we use all stocks in the MSCI World between 07/31/1996 and

06/30/2016. The advantage of this test is that the training data is strictly in the past.

For this test, both models are trained based on weekly data over the past five years and then retrained at the end of every year.[4] Each month, the stocks are sorted into five bins and bought with an equal weight.

The results of the highest quintile are shown Table 4 for the linear and SVR models. The results are shown for each sector and for the combined strategy[5] as well as the equal weighted MSCI World Index. All returns are calculated as total return including gross dividends. Note that the returns are not scaled, but the portfolios are equally weighted between the selected stocks and fully invested.

Table 4. Statistics of the rolling linear and SVR models for each sector, the combined strategy (equal weighted) and the MSCI World Index (equal weighted) based on daily data between 01/01/1996 to 06/30/16

	Materials	Communication	Consumer, Cyclical	Consumer, Non-cyclical	Energy	Financial	Industrials	Technology	Utilities	Combined	MSCI World TR Index (equal weighted)
Linear model											
Annual Mean	0.11	0.15	0.09	0.11	0.12	0.11	0.12	0.13	0.10	0.12	0.09
Annual Vol	0.22	0.20	0.14	0.14	0.28	0.21	0.15	0.26	0.16	0.14	0.15
Eff. Ratio	0.49	0.78	0.64	0.81	0.43	0.52	0.81	0.49	0.66	0.82	0.62
Min Day	-0.09	-0.09	-0.06	-0.06	-0.15	-0.12	-0.06	-0.10	-0.11	-0.06	-0.06
Max Day	0.14	0.13	0.08	0.09	0.14	0.17	0.07	0.30	0.13	0.07	0.07
Skew	0.39	0.36	0.01	-0.07	-0.22	0.81	-0.15	1.43	0.31	-0.23	-0.24
Kurt	11.79	9.34	8.32	10.41	11.50	28.64	7.77	30.97	22.70	9.79	10.03
Max. Drawdown	0.61	0.61	0.43	0.40	0.70	0.70	0.44	0.86	0.29	0.42	0.56
SVR model											
Annual Mean	0.11	0.09	0.10	0.14	0.14	0.10	0.12	0.09	0.10	0.11	0.09
Annual Vol	0.22	0.23	0.19	0.15	0.26	0.20	0.20	0.32	0.16	0.15	0.15
Eff. Ratio	0.49	0.41	0.54	0.93	0.55	0.50	0.63	0.28	0.62	0.72	0.62
Min Day	-0.12	-0.09	-0.08	-0.08	-0.13	-0.09	-0.09	-0.23	-0.07	-0.06	-0.06
Max Day	0.10	0.08	0.12	0.08	0.15	0.10	0.10	0.14	0.08	0.07	0.07
Skew	-0.00	-0.14	0.14	-0.05	-0.18	0.15	-0.01	0.01	-0.23	-0.21	-0.24
Kurt	9.29	7.16	9.88	11.00	8.29	11.51	7.92	12.10	8.07	9.14	10.03
Max. Drawdown	0.62	0.83	0.65	0.54	0.58	0.69	0.57	0.90	0.39	0.52	0.56

The combined strategy equally weights the return of the highest quintile of stocks in each sector and is rebalanced on a monthly basis. This strategy improves the efficiency to 0.82 from 0.62 in the MSCI World Index (equal weighted) with the linear model and to 0.72 with the SVR model. Both models also reduce the skew, kurtosis and maximum drawdown of a global equity investment. The SVR underperforms the linear model in this analysis, but still outperforms the MSCI World TR Index. In particular, the SVR underperforms in the communications sector, but outperforms the linear model in the energy and the non-cyclical consumer sector.

We also analyze the performance of the other quintiles. Selecting the other quintiles result in an efficiency ratio of 0.42 in the lowest quintile with the linear model, which increase monotonously to the highest quintile. Using the SVR model shows similar results.

We analyze the coefficients of the linear model between 2006 and 2015 in more detail (see Appendix C). As expected from our earlier analysis, quality and dividend receive positive factor loadings over time and sectors, while growth (value) receives a negative (positive) loading. Quality receives the most persistent positive loading for stocks from the cyclical consumer sector. Dividend loadings are extremely important in the basic material sector. Growth (Value) is very consistently loaded negatively (positively) for utility stocks.

Interestingly, volatility receives a positive loading in this period and momentum a negative one over the last decade. This is quite the opposite of our first result, but can be explained with the different training periods. The rolling models depend on the very recent five years of data, whereas the first model is trained on the past 10 years. Therefore, bear markets such as in 2000-2002 and 2008 are reflected differently in the machine learning models.

---

[4]With an expanding window in the first five years.

[5]The diversified sector was excluded from this analysis, as only few stocks exist in this category.

Li et al. (2016) and Jegadeesh and Titman (1993) show that low-volatility and positive momentum, respectively, are associated with above average returns. Interestingly, we find the opposite in recent years. Above average returns of the linear model in recent years are rather associated with high-volatility and negative momentum, especially in the communication and technology sector. However, we train the machine learning models on a rolling basis and do not control for collinearity of the input factors. In other words, the models are trained over very different time frames such as bull and bear markets and the different factors might be offsetting each other. For example, high-volatility adds value to the selection of stocks in the communications sector, if the stock shows strong fundamentals otherwise.

6. Discussion

This article applies machine learning to style investing. Style investing has proven to be very successful in recent years and can be improved further with machine learning. Style investing selects stocks not only depending on their market capitalization, but also depending on other factors. We use dividends, momentum, growth, quality and volatility similar to Asness et al. (2015). The fundamental factors are lagged with 6 months to limit data snooping. The proposed machine learning approach learns the efficiency of the factors and weights the stocks in the portfolio accordingly. This approach is easily extendable to other factors and other machine learning methods.

In our first application, we train the machine learning models across time. In other words, the approach neglects the time structure in the training and test data set. Different time points are attributed to each set in a random fashion. As a result, future returns of the training data are overlapping with future returns of the test set. Therefore, the highest quintile of predicted returns might only refer to a certain time period, where the prediction worked well. To measure this overlapping effect of our machine learning models, we test the machine learning models over rolling time periods as well. In this test, the training data is strictly in the past. Interestingly, both analyses show a high outperformance to the equal weighted MSCI World TR Index.[6]

Further research might improve the collinearity of the input factors by applying, for example, a principal component analyses. Arnott et al. (2016) showed that several factor-based strategies, such as momentum, quality or volatility in fact depend on the value factor. We chose to analyze simple factors that are common in the literature instead of building more complex features as input to non-linear models. Another way to build these complex factor combinations would be an application of deep learning techniques that stack several neural networks. The input factors to such a technique could be even more rudimentary data such as balance sheet data which the first layer of a deep network is built from.

Of course, a Sharpe ratio in a long only portfolio above 1 is difficult to achieve in practice. In this article, the Sharpe ratio is based on gross returns, end-of-day data and monthly rebalancing. Trading costs and fees are not included in our analysis. However, Frazzini et al. (2013) show that style trading survives trading costs even at large fund sizes and turnover above 100% per month (in and out) for some strategies. The average turnover in this paper is around 20% per month as the scores are calculated over a large set of data points. Also, this paper only considers the largest and most liquid stocks in global equity markets. Only stocks from the MSCI World index are selected each year.

Most significantly, we find that the input factors contribute differently to the machine learning models in different sectors. Both the linear and SVR model are trained for each economic sector repeatedly. For example, models for financial stocks are driven repeatedly by low volatility, whereas cyclical consumer stocks on momentum. Further research is necessary to analyze the different efficiency of input factor in each economic sector. We find that the contribution of the input factors changes over time and even change their sign in some sectors.

To train the relationship between input factors and future stock performance, we apply both a linear and a Support Vector Regression (SVR) model. The results of the linear model compare nicely to other publications such as Fama and French (2008). Interestingly, the SVR model doubles the efficiency of the linear model in the highest quintile to 1.21. In this article, the SVR model applies standard parameters and a Gaussian Kernel. The standard parameters can be optimized with an in-sample cross-validation procedure to improve results. The choice of a Gaussian Kernel drives the non-linear regression and can be easily extended to other Kernels in the future.

---

[6]In both analyses, the stocks from the highest quintile of multi-factor scores were selected and equal weighted.

References

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross‐section of volatility and expected returns. *The Journal of Finance*, *61*(1), 259-299. http://dx.doi.org/10.1111/j.1540-6261.2006.00836.x

Arnott, R., Beck, N., Kalesnik, V., & West, J. (2016). How Can 'Smart Beta' Go Horribly Wrong? Fundamentals.

Asness, C., Ilmanen, A., Israel, R., & Moskowitz, T. (2015). Investing with style. *Journal of Investment Management, 13*(1), 27-63.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297. http://dx.doi.org/10.1007/BF00994018

Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in Statistics* (pp. 569-593). Springer New York. http://dx.doi.org/10.1007/978-1-4612-4380-9_41

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance, 47*(2), 427-465. http://dx.doi.org/10.1111/j.1540-6261.1992.tb04398.x

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics, 33*(1), 3-56. http://dx.doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (2008). Dissecting anomalies. *The Journal of Finance, 63*(4), 1653-1678. http://dx.doi.org/10.1111/j.1540-6261.2008.01371.x

Fama, E. F., & French, K. R. (2015). International tests of a five-factor asset pricing model. *Fama-Miller Working Paper*. http://dx.doi.org/10.2139/ssrn.2622782

Frazzini, A., Israel, R., & Moskowitz, T. J. (2012). Trading costs of asset pricing anomalies. *Fama-Miller Working Paper*, 14-05.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *The Annals of Statistics*, 927-953. http://dx.doi.org/10.1214/aos/1176350933

Hanna, J. D., & Ready, M. J. (2005). Profitable predictability in the cross section of stock returns. *Journal of Financial Economics, 78*(3), 463-505. http://dx.doi.org/10.1016/j.jfineco.2004.12.004

Haugen, R. A., & Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics, 41*(3), 401-439. http://dx.doi.org/10.1016/0304-405X(95)00868-F

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance, 48*(1), 65-91. http://dx.doi.org/10.1111/j.1540-6261.1993.tb04702.x

Lewellen, J. (2014). The cross section of expected stock returns. Forthcoming in *Critical Finance Review*. http://dx.doi.org/10.2139/ssrn.2511246

Li, X., Sullivan, R. N., & Garcia-Feijóo, L. (2016). The Low-Volatility Anomaly: Market Evidence on Systematic Risk vs. Mispricing. *Financial Analysts Journal, 72*(1), 36-47. http://dx.doi.org/10.2469/faj.v72.n1.6

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics, 108*(1), 1-28. http://dx.doi.org/10.1016/j.jfineco.2013.01.003

Appendix A

Cross-validation and Bootstrapping

For our first analysis, we partition the data into training and test sets to analyze robustness and out-of-sample performance of the models. The k-fold cross-validation procedure divides the observations into k disjoint subsamples (or folds) from n data points, chosen randomly but with approximately equal size. We chose k as 5, to balance training on 80% of the data and test set on 20% of the data. The 20% of data points in the test set are then labelled by the machine learning models with the predicted future return and sorted into quintiles.

5-fold cross-validation leads to five different test sets to test the robustness of the model. On each test set our hypothesis is tested. To boost the number of tests, we also perform a bootstrapping technique to increase the test set to 1000 (see Efron, 1979). For bootstrapping, we randomly sample the training and test set with replacement. This technique allows our hypothesis to be tested for 1000 times and generate a more detailed estimate. Both methods operate on 20% of the data for testing and strictly separate training data. Since the 5-fold cross-validation is limited to 5 different test sets, we chose bootstrapping to evaluate the bias due to sampling and test the robustness of our model parameters.

The two-sample t-test is performed separately for the two different time periods and for the different sectors. Also, the test is performed five times, for the 5-fold cross-validation and 1000 times for the bootstrapped method. The Table A1 shows the percent of cross-validations the H0 is rejected for both cross-validation methods. The table also shows the number of data points in the test sample in each quintile.[7]

Table A1. Significance test for the linear model

	07/31/1996 to 07/31/2006			07/31/2006 to 06/30/2016		
	5-fold cross-validation	Bootstrapped	Sample size	5-fold cross-validation	Bootstrapped	Sample size
Basic Materials	100%	98%	425	80%	79%	356
Communications	100%	100%	317	80%	76%	233
Consumer, Cyclical	100%	99%	712	100%	99%	551
Consumer, Non-cyclical	100%	100%	721	100%	96%	571
Diversified	60%	60%	76	30%	27%	58
Energy	100%	100%	188	70%	69%	186
Financial	100%	100%	838	100%	98%	661
Industrials	100%	100%	811	100%	97%	721
Technology	50%	46%	156	30%	34%	101
Utilities	90%	94%	278	80%	79%	181

Note. Percent of cross-validations the H0 is rejected at the 5% level and the number of data points in the test sample per quintile for the two different periods. The percentage is reported both for the 5-fold cross-validations and the bootstrapped method.

The results for the 5-fold cross-validation and the bootstrapped method are very similar. For both methods, the table shows that the linear model is very successful and stable in separating the best performing from the worst performing quintiles of stocks. For example in the utilities sector during the period 1996 to 2006, in 94 out of 100 bootstrapped cross-validations the selected stocks in the highest bucket were performing significantly better than the selected stocks in the lowest bucket over the next month. Since the 5-fold cross-validation is a random selection, it returns different results for each repetition. The bootstrapping method is giving a more detailed and replicable measure of the test.

In both periods, the cross-validation only drops in the diversified and technology sectors. These are also the two sectors that have the least data to fill the different buckets. Especially for the diversified sector, the data set only comprises 58 data points for each of the five buckets in the period 07/31/2006 to 06/30/2016.

We perform the same test for the Support Vector Regression (SVR) model (see Table A2). The results show that the SVR model is outperforming the linear model by far. The SVR always separates the highest and lowest quintile significantly across sectors, time and cross-validation techniques for the same data set as the linear model.

---

[7]The sample size is defined as 20% of the total data points in a sector in each time period and then divided by 10 buckets. For example, the total number of data points in Basic Materials was 21.250. Out of these 17.000 were used for training of the model and 4250 were used for testing. For the test sample, the model predicted future returns and these were then aggregated to buckets of 425 data points.

Table A2. Significance test for the Support Vector Regression (SVR) model

	07/31/1996 to 07/31/2006			07/31/2006 to 06/30/2016		
	5-fold cross-validation	Bootstrapped	Sample size	5-fold cross-validation	Bootstrapped	Sample size
Basic Materials	100%	100%	425	100%	100%	356
Communications	100%	100%	317	100%	100%	233
Consumer, Cyclical	100%	100%	712	100%	100%	551
Consumer, Non-cyclical	100%	100%	721	100%	100%	571
Diversified	100%	100%	76	100%	100%	58
Energy	100%	100%	188	100%	100%	186
Financial	100%	100%	838	100%	100%	661
Industrials	100%	100%	811	100%	100%	721
Technology	100%	100%	156	100%	100%	101
Utilities	100%	100%	278	100%	100%	181

Note. Percent of cross-validations the H0 is rejected at the 5% level and the number of data points in the test sample per quintile for the two different periods. The percentage is reported both for the 5-fold cross-validations and the bootstrapped method.

Appendix B

Feature Analysis

The Figure B1 below shows example results for the current linear and Support Vector Regression (SVR) models in the utility sector. The models are trained with data prior to 01/01/2016. The figure shows the response of the linear and SVR model to the different input factors. The response of the models is measured by the predicted future return.

In a linear model, the score increases or decreases linearly with higher input values. In the example below, a higher dividend yield (divYield) creates a higher predicted future return. The response of the SVR model is more complex. The SVR model is converging into two main prediction areas, one for extreme values in the input factors and 'normal' values. For extreme values, the SVR model returns the average return of the stocks in the sector over the trained period.
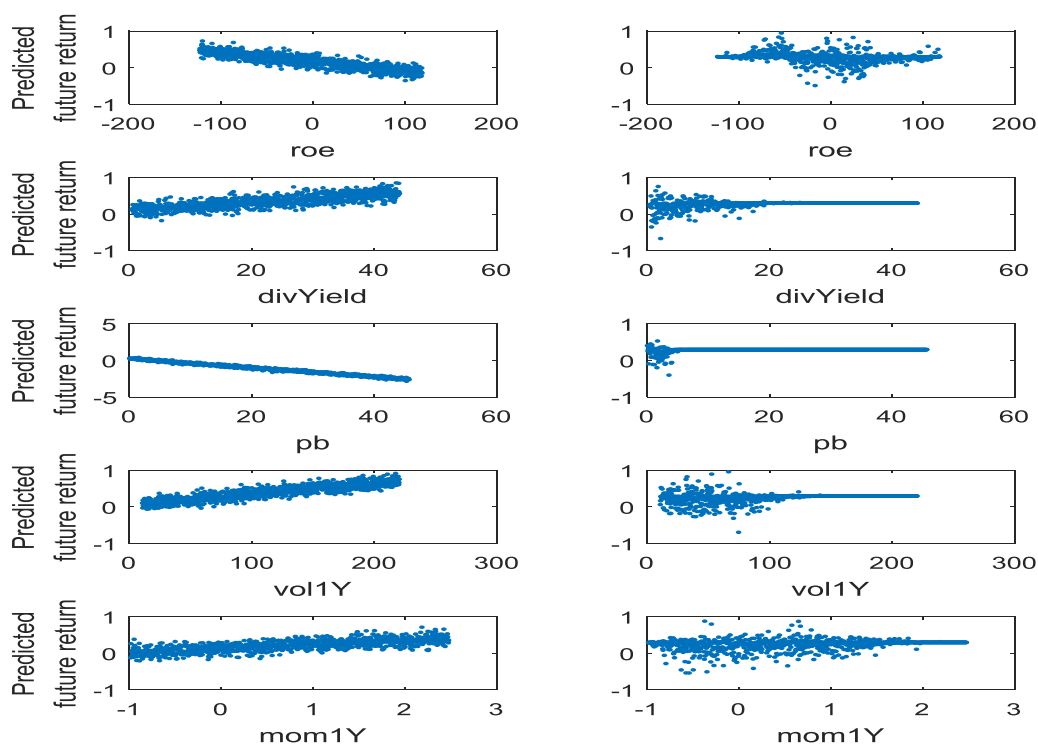

Figure B1. Example response of the linear model (left column) and the SVR model (right column) in the "utility" sector

Note. The model is trained on data between 01/01/2011 and 01/01/2016 (4 years of factor values and 1 year of future return). Response is shown as the relationship between the predicted future return to different values of the five different factors.

Appendix C

Coefficents of the Rolling Linear Model

Table C1. Coefficients of the linear model for each year between 2006 and 2015

		29-Dec-2006	31-Dec-2007	31-Dec-2008	31-Dec-2009	31-Dec-2010	30-Dec-2011	31-Dec-2012	31-Dec-2013	31-Dec-2014	31-Dec-2015
Quality	Mat	-0.02%	0.07%	0.05%	0.26%	0.17%	0.00%	0.12%	0.02%	-0.06%	0.04%
	Com	-0.01%	0.00%	-0.01%	0.01%	0.09%	0.05%	-0.02%	-0.02%	-0.02%	-0.01%
	Cycl	0.09%	0.06%	-0.03%	0.03%	0.06%	0.08%	0.10%	0.10%	0.12%	0.10%
	NCycl	-0.03%	-0.01%	0.00%	0.00%	0.00%	0.01%	0.02%	0.02%	0.02%	0.01%
	Ergy	-0.26%	-0.32%	-0.22%	0.00%	0.07%	0.00%	-0.02%	0.05%	0.13%	0.36%
	Fincl	-0.08%	-0.05%	-0.24%	-0.04%	0.23%	0.27%	0.29%	0.23%	0.02%	-0.02%
	Indu	0.10%	0.06%	-0.02%	-0.05%	-0.03%	-0.06%	0.01%	0.05%	-0.06%	-0.02%
	Tech	-0.22%	-0.11%	-0.08%	-0.01%	0.10%	0.14%	0.11%	0.05%	0.01%	-0.02%
	Util	0.08%	0.18%	0.06%	0.12%	0.04%	-0.09%	0.00%	-0.10%	0.09%	0.43%
Dividend	Mat	0.36%	0.78%	0.78%	3.75%	6.04%	5.85%	5.99%	3.84%	2.01%	0.66%
	Com	1.53%	1.11%	0.22%	0.05%	0.05%	0.03%	0.06%	0.13%	-0.92%	-0.84%
	Cycl	1.14%	1.81%	2.16%	2.53%	3.04%	2.55%	2.18%	0.57%	0.13%	0.09%
	NCycl	0.01%	0.01%	0.00%	0.00%	-0.20%	-0.21%	-0.08%	-0.05%	-0.01%	-0.06%
	Ergy	1.15%	0.68%	0.82%	0.81%	1.66%	2.08%	2.57%	1.28%	0.02%	-0.56%
	Fincl	1.21%	0.42%	0.59%	0.78%	1.73%	1.66%	1.70%	1.82%	0.17%	0.19%
	Indu	2.13%	1.84%	1.55%	2.20%	3.17%	2.94%	1.53%	0.68%	0.03%	-0.12%
	Tech	-0.15%	-0.36%	-0.39%	0.46%	0.29%	0.30%	0.63%	0.60%	0.69%	0.30%
	Util	1.10%	0.47%	1.36%	0.66%	0.97%	1.38%	1.14%	1.96%	2.50%	1.42%
Growth	Mat	-0.01%	-0.41%	-0.01%	-0.67%	0.03%	-0.45%	-0.49%	0.91%	0.92%	-0.05%
	Com	-0.11%	-0.08%	-0.04%	-0.02%	-0.20%	-0.07%	0.07%	0.09%	0.12%	0.10%
	Cycl	-0.82%	-0.57%	-0.05%	-0.04%	-0.04%	-0.04%	0.14%	0.05%	-0.03%	-0.05%
	NCycl	0.00%	-0.03%	-0.04%	-0.06%	-0.08%	-0.13%	-0.08%	-0.01%	-0.01%	-0.01%
	Ergy	0.01%	0.01%	0.02%	0.02%	1.24%	-0.39%	-0.81%	0.40%	-0.41%	-0.96%
	Fincl	-0.52%	-0.07%	0.02%	-0.04%	-0.12%	-0.97%	-0.42%	0.23%	0.56%	0.56%
	Indu	0.01%	0.86%	-0.50%	-0.77%	-0.11%	-0.04%	-0.01%	0.01%	0.12%	0.10%
	Tech	0.57%	0.33%	-0.10%	-0.13%	-0.19%	-0.16%	-0.20%	-0.12%	-0.03%	0.04%
	Util	-1.24%	-0.65%	-1.78%	-3.29%	-3.34%	-5.21%	-4.97%	0.33%	1.07%	-4.66%
Volatility	Mat	0.85%	0.87%	-0.26%	0.34%	0.81%	1.03%	1.04%	0.58%	-0.15%	-0.34%
	Com	0.87%	0.67%	0.01%	0.40%	1.08%	1.23%	0.98%	0.72%	0.86%	0.99%
	Cycl	0.58%	0.78%	0.10%	0.23%	0.86%	1.08%	0.83%	0.51%	0.34%	0.35%
	NCycl	0.41%	0.41%	-0.36%	0.22%	0.68%	0.81%	0.66%	0.29%	0.16%	0.20%
	Ergy	-0.15%	-0.14%	-0.14%	0.48%	0.66%	0.76%	0.79%	0.41%	0.04%	0.26%
	Fincl	0.31%	0.36%	-0.59%	0.01%	0.59%	0.67%	0.53%	0.40%	0.37%	0.66%
	Indu	0.51%	0.50%	-0.28%	0.03%	0.64%	0.92%	0.91%	0.53%	0.19%	0.22%
	Tech	0.71%	0.51%	-0.21%	0.52%	1.14%	1.29%	1.08%	0.47%	-0.15%	-0.13%
	Util	0.55%	0.46%	0.57%	-0.32%	0.03%	0.23%	0.18%	0.14%	0.20%	0.20%
Momentum	Mat	3.02%	1.24%	20.78%	4.32%	-2.09%	-1.80%	-5.08%	2.42%	10.53%	-0.64%
	Com	-5.98%	-3.40%	9.61%	7.76%	-10.72%	-7.44%	-7.73%	-6.68%	-8.32%	-7.25%
	Cycl	-2.81%	-4.55%	7.68%	-2.29%	-8.01%	-3.29%	-4.23%	-2.95%	2.80%	0.15%
	NCycl	1.74%	8.21%	16.82%	1.98%	-11.40%	-11.98%	-10.78%	-5.82%	2.83%	-0.56%
	Ergy	16.74%	7.01%	6.90%	-8.76%	-26.7%	-13.38%	-8.84%	-0.43%	4.92%	3.22%
	Fincl	3.51%	3.14%	28.38%	16.42%	-2.82%	-10.93%	-12.83%	-14.58%	-7.50%	-10.32%
	Indu	1.59%	3.63%	11.43%	3.24%	-8.08%	-9.26%	-9.31%	-7.07%	-0.46%	-5.44%
	Tech	-7.99%	-4.85%	13.31%	-7.34%	-18.48%	-10.17%	-11.21%	-6.17%	3.86%	-3.17%
	Util	16.86%	7.88%	5.65%	10.22%	-2.52%	-7.14%	-2.21%	-0.54%	3.53%	-7.72%

Note. The coefficients are shown for each input factor and each sector separately.