Stock returns and the cross-section of characteristics: a tree-based approach

Guillaume Coqueret* Tony Guida†

April 27, 2018

Abstract

We build regression trees to determine which firm characteristics are most likely to drive future returns. Out of 30 attributes, those related to momentum appear to have, by far, the most marked impact. This prominence is verified at the sector level as well. The second order effects are propelled by volatility and liquidity variables. Finally, we show that a realistic portfolio strategy based on the short-term RSI characteristic outperforms the naive 1/N portfolio by 2.4% per annum, once the transaction costs have been taken into account. One possible explanation for these higher returns is the immunity of the strategy to the momentum crash phenomenon.

Keywords: Regression trees, Cross-section of stock returns, Firm characteristics, Portfolio choice
JEL Codes: G12, G11, C44, C55

1 Introduction

The complexity and multiplicity of drivers of stock returns are such that they cannot be rendered by simple asset pricing models (e.g. the CAPM and its offspring). These latter models are appealing because they provide insightful relationships, but they fail to reproduce stylized facts observed empirically. Beyond the critique of Roll (1977), the most striking of these failures is highlighted in the seminal work of Fama and French (1992) who popularize the so-called size and value anomalies.[1] Firms with small market capitalization (*resp.* high book-to-market ratios) are shown to experience higher returns than their large capitalization (*resp.* low book-to-market) counterparts. This spread in performance is in contradiction with the CAPM, hence the term "anomaly".

The literature on asset pricing anomalies is so vast that it has own meta-studies (Subrahmanyam (2010), Goyal (2012), Green et al. (2013), and Harvey et al. (2015)). Recent contributions shed light on specifics aspects: investor sentiment and limits to arbitrage (Jacobs (2015)), predictability (McLean and Pontiff (2016)), Fama-McBeth regressions (Green et al. (2017)) and portfolio selection (Brandt et al. (2009), Hjalmarsson and Manchev (2012), Ammann et al. (2016), Moritz and Zimmermann (2016) and DeMiguel et al. (2017)).

These so-called "anomalies" have generated a strand of literature (started by Fama and French (1993)) dedicated to pervasive factors underlying the cross-section of stock returns. The contributions to this field are so numerous they cannot be reviewed here and we refer to Green et al. (2013), and Harvey et al. (2015) for exhaustive surveys. Nonetheless, evidence brought forward by Daniel and Titman (1997) and

*Montpellier Business School, 2300 Avenue des Moulins, 34080 Montpellier, France (corresponding author)
†UniversitŐ Savoie Mont Blanc

[1] Their results are in line with earlier findings (e.g. Banz (1981) or Graham and Dodd (1934)), but it is clearly after 1992 that the research on these anomalies has gained traction.

Daniel et al. (2001) indicates that stock returns are more likely to be propelled by firms' characteristics than by priced factors.[2] Inspired by these conclusions, our work intends to apply a cornerstone of data science[3] to 30 firm characteristics so as to shed light on which are the ones that matter the most for pricing purposes.

With the emergence of Big Data, Machine Learning techniques have become appealing to explain or forecast stock prices and returns. For instance, Neural Networks have been employed as early as the mid-1990s (Bansal and Viswanathan (1993)), but they are now a commonplace tool (Kara et al. (2011), Krauss et al. (2016)). Trees, too, have been used, both in the past (Sorensen et al. (2000)) and recently (Moritz and Zimmermann (2016)), but only for portfolio construction purposes.[4] The protocols put forward in these studies do not fully reveal which characteristics (or combinations thereof) are pregnant to explain stock returns. The main purpose of this paper is to partially fill this void.

In this article, we perform regression trees on a sample that comprises 30 classical characteristics. A key improvement of trees is that they take into account conditional impact of characteristics, given the value of other characteristics. This is a feature that linear models (estimated through regressions) fail to take into account. Our main finding is that the attributes which most often lie close to the root of the trees are all related to momentum, especially the 3 month Relative Strength Index. This highlights the importance of both technical indicators and past performance in the pricing of stocks. It is likely that these conclusions have behavioral foundations and they can be put into perspective along recent findings. Technical analysis has been found to explain and forecast returns both at the aggregate level (Neely et al. (2014), Taylor (2014) and Lu et al. (2015)) and in the cross-section of assets (Han et al. (2013) and Han et al. (2016)). Similarly, recent performance (momentum) is ubiquitous in the literature since the empirical work of Jegadeesh and Titman (1993).[5] Our results are yet another confirmation of these effects.

The remainder of this paper is structured as follows. In Section 2, we present our database. In Section 3, we outline the principles of the method and discuss our baseline results. Section 4 is dedicated to extensions (chronological analysis, sectorial focus and portfolio choice implications). Finally, we conclude.

2 Data

We start by describing our dataset. Our sample consists of large US stocks originating from the Capital IQ database. The sample starts in January 2002 and ends in June 2016. The number of firms in the sample oscillates between 305 and 599, with an average value of 454. The variables (i.e characteristics) are described in Table 3 in the Appendix. We have split them into four categories: accounting, momentum, volatility and liquidity. They are recorded at the monthly frequency. In the Appendix, in Table 4, we provide the descriptive statistics of the characteristics over the whole sample.

[2] Recently, Suh et al. (2014) have proposed a new PCA-based method which accounts more effectively for time-series and cross-sectional variations in returns.

[3] Decision trees go back at least to the 1960s: e.g., Morgan and Sonquist (1963).

[4] Other more exotic techniques include fuzzy regression (Kocadagli (2013)) and online sentiment (Kim and Kim (2014)).

[5] Other references on the topic include (the list is far from exhaustive): Jegadeesh and Titman (2001), Menkhoff (2011), Asness et al. (2013).

Mechanically, the dynamic shifts of the characteristics impose that they be normalized at each date. Indeed, we seek to link the impact of the *relative* (across stocks, not across time) values of the characteristics on future returns. This is a standard procedure (see Brandt et al. (2009), for instance). There are many possible normalizations. We choose to work with the empirical cumulative function (ecdf) of the characteristics so that their values lies between zero and one: at each date, the smallest transformed characteristic is equal to zero and the largest one is equal to one. This choice is very convenient because all variables then have the same range and can be easily compared. Also, in the trees, this choice makes it easy to interpret the splits: 0.5 corresponds to the median, 0.25 to the first quartile, etc.

All of the variables in the study are standard in the financial economics and accounting literature. In anticipation to our results, we feel it is useful to provide details on the RSI characteristic. The Relative Strength Index (RSI) is a technical indicator computed as follows:

$$RSI(n) = 100 \times \frac{H(n)}{H(n)+B(n)},$$

where

$$H(n) = \frac{2}{n+1}I(n) + \frac{n-1}{n+1}H(n-1),$$
$$B(n) = \frac{2}{n+1}D(n) + \frac{n-1}{n+1}B(n-1),$$

and $I(n)$ (*resp.* $D(n)$) are the latest increase (*resp.* absolute decrease) in prices over the last period (week, in our sample). n is the number of observations required to compute the moving averages $H(n)$ and $B(n)$. It is equal to 13 for the 3 month RSI and to 52 for the 12 month RSI. When a stock goes up regularly, its RSI increases to 100 and when it goes down, it decreases to zero. As such, the RSI can be viewed as an indicator of momentum.

Given all of the characteristics, the aim is to build decision trees with future returns as dependent variable. It is the purpose of the following section.

3 Principle and results

3.1 Regression trees

Regression trees are a top-down classification tool. They iteratively split a sample into coherent clusters. We refer to the monograph Friedman et al. (2009) for formal details on the topic and we briefly present the principle and functioning of the trees below.

We assume that the dataset comprises $K+1$ characteristics. The first K items characterize the firms (market capitalization, book-to-market ratio, past returns, earnings, volatility, etc.). They can be viewed as independent variables. Item number $K+1$ is a future (1 month, 6 month, 1 year) total return of the corresponding firm and will serve as dependent variable. The database can be represented as a $T \times (K+1)$ matrix $\boldsymbol{C} = c_{i,j}$. The columns are the characteristics and each line corresponds to a pair consisting of one date and one company.

At the root of the tree, the optimal split for characteristic j is such that the two clusters formed according to this variable have the smallest (total) variability in future returns: $s_j^* = \underset{s}{\operatorname{argmin}}\ V_j^s$, with

$$V_j^s = \sum_{i=1}^{T} 1_{(c_{i,j}>s)}(c_{i,K+1} - \mu_j^{s+})^2 + \sum_{i=1}^{T} 1_{(c_{i,j}\le s)}(c_{i,K+1} - \mu_j^{s-})^2, \tag{1}$$

where

$$\mu_j^{s+} = \frac{\sum_{i=1}^{T} 1_{(c_{i,j}>s)}c_{i,K+1}}{\sum_{i=1}^{T} 1_{(c_{i,j}>s)}} \text{ and } \mu_j^{s-} = \frac{\sum_{i=1}^{T} 1_{(c_{i,j}\le s)}c_{i,K+1}}{\sum_{i=1}^{T} 1_{(c_{i,j}\le s)}}$$

are the future return averages related to the two clusters. $1_{(\cdot)}$ is the indicator operator: $1_{(x)}$ is equal to 1 if x is true and to zero otherwise.

This step can be processed for all characteristics, and the final split is performed for the one that reaches the smallest (scaled) variance $V_j^{s_j^*}$. The sample is thus split into two and new partitions can be iterated on the two new subsets. The dataset is then subdivided into increasingly homogeneous clusters of future returns. Each cluster depends on a particular succession of splits. The relationship between future returns and firms' characteristics is thus highly non-linear (as opposed to regression for instance). The question of the optimal number of leaves (i.e. final clusters) is technical (e.g. Mingers (1989)) and somewhat arbitrary. In order to make the trees easy to read, we will restrict our trees to 10 leaves at most. This is no true limitation because the early splits are those which are most important: they show which characteristics are the key drivers of the future returns.

3.2 Results

In Figure 1, we present the tree when the dependent variable is the future 1 month, 6 month or 12 month returns. In the trees, the software determines which splits are optimal (i.e. which ones are those which reduce the dispersion in (1) the most). Therefore, the structure of the trees is never the same. The trees can be read in the following way. If we look at the first tree of Figure 1, the root is at the top. The sample is split according to the RSI 3M variable: to the left below the root are gathered the occurrences where the RSI 3M is lower than the 49.1% quantile (in the cross-section) and all remaining occurrences are grouped in the cluster to the right of the root. Then if we consider the cluster on the left, it is again split in two according to the same variable, and below to its left are the cases when the RSI 3M is lower than the 16.7% quantile. This cluster is then split depending on share turnover.

In red, we display the average future returns of each cluster. For instance, the average 1M forward return is equal to 0.007 over the whole sample, but for firms with RSI 3M lower (*resp.* higher) than the 0.491 quantile, this average falls to -0.011 (*resp.* increases to +0.025). The secondary splits confirm the strong discriminative power of the RSI 3M variable: the average future return is -0.028 below the 0.167 quantile and +0.043 above the 0.831 quantile.

In the first tree, a striking feature is that only two variables are considered relevant by the algorithm: the 3-month relative strength index and the share turnover. This pattern is also salient when 6 month returns are considered, except that volatility emerges as a third-order splitting criterion. With regard to long term (12 month) returns, many characteristics appear on the right-hand side of the tree, but the first (and major) split is still performed according to the RSI 3M attribute.

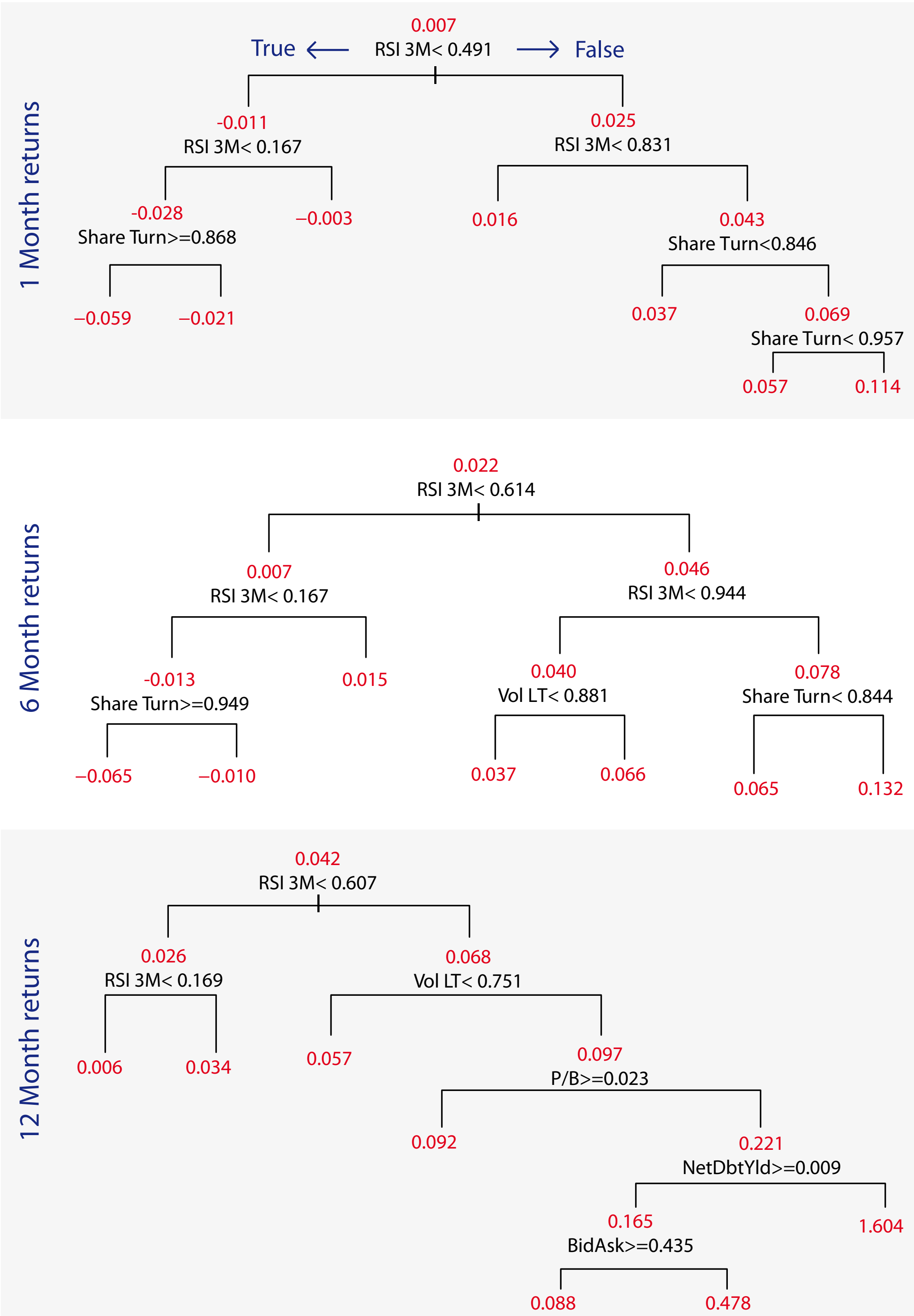

Figure 1: **Regression trees with 1, 6 and 12 month future returns as dependent variable**. The splitting criteria are indicated in black and the average future 1, 6 and 12 month returns are displayed in red for each node. The cluster corresponding to the criteria lies to the left of the node and the opposite cluster is located to its right. The leaves are ordered so that returns display an overall increasing pattern. The trees are pruned to improve readability.

These findings are in line with recent conclusions in the asset pricing and portfolio choice literature. For example, Neely et al. (2014) show that technical indicators can significantly forecast the equity premium and Han et al. (2016) introduce a trend factor which captures price trends at different horizons and also performs well at explaining the cross-section of stock returns.[6] The RSI can be viewed as a trend indicator, hence it appears that our results confirm, via a tree-based approach, the prominence of momentum effects among the drivers of stock returns.

3.3 Comparison with a regression

We now aim to put our results into perspective by comparing them with the insights provided by a simple (linear) regression. With the notations of Section 3.1, we estimate the following equation:

$$c_{i,K+1} = \alpha + \sum_{k=1}^{K} \beta_k c_{i,k} + \epsilon_i, \tag{2}$$

where on the l.h.s., the dependent variable is the future return and on the r.h.s. lie all of the firms' characteristics. In Table 1, we provide the estimates as well as the corresponding t-stats and their significance levels.

Even if many variables appear to have a significant predictive power over future returns, the t-stats of the RSI variables clearly indicate that their impact is decisive. Now, let us focus on one month forward returns as dependent variable. In the upper tree of Figure 1, two variables stand out: the RSI 3M and the share turnover. With regard to the simple regression, the RSI 3M emerges as a major driver of one month returns, but not the share turnover (it ranks fourteenth in absolute t-stat ranking). The reason for this is rooted in the non-linear peculiarity of the tree: share turnover has opposite effects depending on the level of RSI 3M. If the RSI 3M is high, then liquidity (for which share turnover is a proxy) is beneficial to future returns; but when the RSI 3M is low, then liquidity becomes hurtful. Because of this contradictory conditional impact of share turnover, the regression fails to identify it as a key characteristic because the positive effects of liquidity are aggregated with (and hence mitigated by) the negative effects. The exact same conclusions hold for 6 month forward returns.

Turning to one-year returns, we can again acknowledge discrepancies in outcome between the tree and the regression. In the tree, the second most important variable is the realized long-term volatility, whereas it ranks sixteenth in terms of absolute t-stat in the regression. Here again, the regression cannot identify it as a strong driver of future returns because long-term volatility matters only when the RSI 3M is high.

Overall, although both methods agree on the most important predictor of future returns (the RSI 3M), the second-order effects are clearly different. While the trees allow for conditional effects, the regression can solely compile additional variables.

[6] In a similar perspective, Moritz and Zimmermann (2016) find that among one hundred attributes including many accounting figures, past performance (i.e., recent returns) is overwhelmingly represented in the construction of their tree-based portfolios.

	1M future returns			6M future returns			12M future returns		
Variable	Coef.	t-stat		Coef.	t-stat		Coef.	t-stat	
(Intercept)	-0.034	-8.637	(***)	0.000	0.033		0.020	2.552	(*)
DivYld	-0.022	-8.024	(***)	-0.035	-8.348	(***)	-0.035	-6.378	(***)
BuyBackYld	-0.002	-0.688		-0.012	-3.235	(**)	-0.030	-5.916	(***)
NetDbtYld	0.011	3.227	(**)	0.013	2.694	(**)	0.007	1.131	
ShareHldrYld	-0.008	-1.940	(.)	-0.006	-0.935		0.004	0.457	
FCFYld	0.002	1.157		0.007	2.585	(**)	0.005	1.601	
PyoutRatio	0.012	4.229	(***)	0.022	5.122	(***)	0.008	1.403	
GrProfitGrowth	-0.002	-1.297		0.004	1.541		0.005	1.713	(.)
PB	0.015	4.201	(***)	0.009	1.744	(.)	-0.002	-0.237	
ROE	-0.009	-2.843	(**)	-0.010	-2.028	(*)	-0.024	-3.657	(***)
ROC	-0.003	-0.897		-0.009	-1.815	(.)	0.000	-0.050	
DebtEquity	0.006	2.373	(*)	0.012	2.987	(**)	0.021	3.932	(***)
NetDebt	-0.005	-1.996	(*)	-0.014	-3.422	(***)	-0.024	-4.532	(***)
MKTCAP	0.002	0.546		0.026	4.033	(***)	0.073	8.570	(***)
BidAsk	-0.012	-8.071	(***)	-0.007	-2.913	(**)	0.004	1.440	
ADV20	-0.059	-7.367	(***)	-0.071	-5.957	(***)	-0.082	-5.152	(***)
ADV60	0.019	1.753	(.)	0.047	2.967	(**)	0.091	4.286	(***)
ADV12M	0.046	5.707	(***)	0.007	0.566		-0.068	-4.248	(***)
MOM 12M	-0.024	-11.713	(***)	-0.014	-4.540	(***)	-0.001	-0.318	
MOM 6M	-0.008	-4.288	(***)	-0.005	-1.679	(.)	-0.006	-1.601	
Vol LT	0.000	-0.049		-0.013	-2.727	(**)	-0.015	-2.410	(*)
Vol ST	0.001	0.292		0.014	2.996	(**)	0.011	1.720	(.)
RSI 12M	0.030	19.210	(***)	0.023	10.016	(***)	0.019	6.220	(***)
RSI 3M	0.097	69.983	(***)	0.075	36.393	(***)	0.079	28.876	(***)
Op Prt Margin	0.007	1.754	(.)	0.004	0.703		0.014	1.696	(.)
Net Prt Margin	-0.007	-1.969	(*)	0.002	0.343		0.009	1.344	
Share Turn	-0.005	-2.010	(*)	0.007	2.180	(*)	0.023	4.910	(***)
EV/EBITDA	0.008	3.042	(**)	-0.010	-2.372	(*)	-0.015	-2.721	(**)
Cash Conv.	0.002	1.016		-0.004	-1.535		-0.007	-1.871	(.)
PE	-0.003	-1.251		-0.007	-1.841	(.)	0.005	0.999	
Psales	-0.004	-0.811		-0.018	-2.620	(**)	-0.045	-4.918	(***)

Table 1: **Linear regression results**. Future returns are regressed against all other characteristics. The significance levels for the t-stats are: (***)<0.001<(**)<0.01<(*)0.05<(.)<0.1.

4 Extensions

In this section, we extend our base-case results in three directions. First, we perform a chronological partition of our findings to unveil whether characteristics' importance is stable through time. Second, we subdivide our sample into industries and check whether results vary from one sector to another. Finally, we build portfolios based on the RSI variable and assess the potential profits for a stock-market investor.

4.1 Characteristics' importance across time

The results of Section 3 were obtained on the whole sample, but it is likely that the characteristics of firms have changing impacts, depending on market or economic conditions (e.g., Peltomäki and Äijö (2015)). Accordingly, we seek to quantify the variations in importance (within the trees) for each characteristic. Unfortunately, plotting 15 trees (there are 15 calendar years in our sample) would make them hard to read and interpret succinctly.

We thus resort to the variable importance metric introduced in Atkinson and Therneau (2015)[7]. This metric, when normalized, gives, on a scale form zero to one, the relative importance of each variable in the tree. Variables close to the root logically benefit from a larger weight. In Figure 2, we display the importance metric (normalized so as to equate one each year) on stacked bar plots.

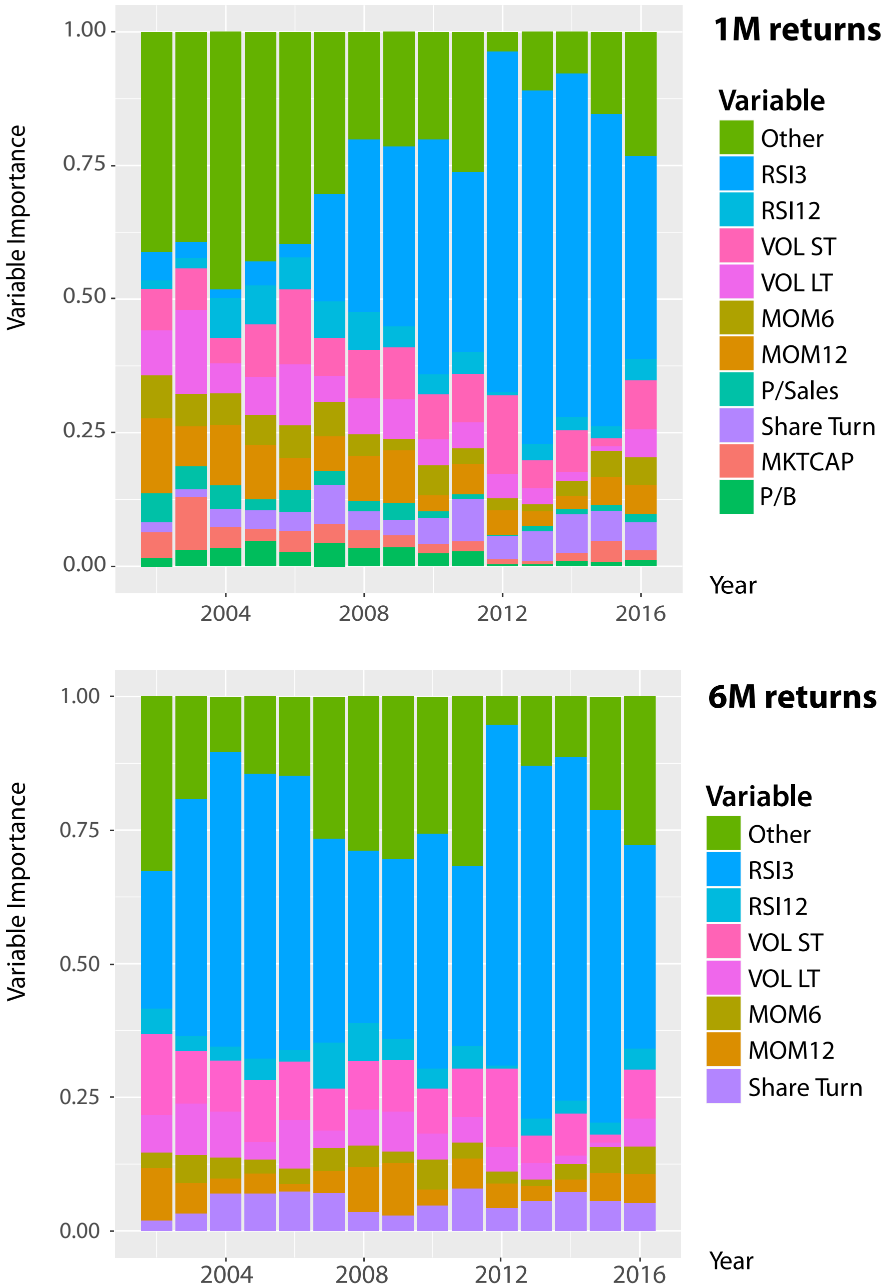

Figure 2: **Variable importance across time**. Each year, the metric is computed and then normalized so that the sum equals one. Only the 10 most represented variables (highest averages over the 15 years) are shown in detail. The dependent variable in the regression trees is the 1 month future return.

[7]Loosely speaking, it is equal to the sum of the goodness of split measures for each split for which it was the primary variable, plus an adjustment term for all splits in which it was a surrogate. For more details, we refer to Atkinson and Therneau (2015). We make use of the rpart package with precision (cp) parameter equal to 0.002, which ensures a suitable accuracy.

To ease readability, we have aggregated the minor variables (those with the smallest average importance over the 15 years) into a "*Other*" category. We provide the results for one- and six-month forward returns. In the top graph, we see that the RSI 3M variable has a relatively modest importance in the early years of the sample but gains traction progressively and clearly dominates from 2008 onwards. If we add the RSI and momentum variables, which are all related to trend indicators, the aggregate importance oscillates between 20% (in 2003) and 73% (in 2013) with an average value of 47%. The second prominent family of indicators are the realized volatilities (in pink on the graphs). However, with an average importance metric of 15%, they lag far behind the trend characteristics. It is nonetheless interesting to mention that in Table 1, the two volatility proxies are deemed irrelevant by the regression (for one-month returns). This is a point for which the two approaches strongly disagree.

Similar conclusions can be drawn for the second graph where in fact we see that the importance of the RSI 3M is high even at the beginning of our sample. In unreported results, we performed the same analysis for one year forward returns and the outcome was far more balanced: no particular characteristic (or group thereof) stands out and emerges as an essential driver of future one year returns. This is logical, given the numerous variables in the corresponding tree (lower part of Figure 1).

4.2 Sector splits

In this section, we perform the same analysis at the sector level to test whether or not specific characteristics emerge as drivers of future returns when firms have somewhat homogeneous activities. We build the corresponding trees for the three sectors (out of the 11 major Morgan Stanley Capital International (MSCI) sectors) for which more than 10,000 observations were available: Consumer Discretionary, Financials and Information Technology.[8]

In Figure 3, we plot the corresponding trees with one-month forward returns as dependent variable. A striking observation is that, for all three sectors, the first split is again determined by the RSI 3M variable. For financial and IT firms, the secondary splits (volatility or free cash-flow yield) are irrelevant because they concern very high (close to one) or very low (close to zero) quantiles, which means that they correspond to only a handful of cases. In short, this means that the algorithm can only marginally improve the bulk of the clusters. The fact that only the primary split matters is yet another confirmation of the prevalence of the RSI 3M variable over the other characteristics. Also, this shows that our main conclusion is robust and not sector-dependent.

[8] These three sectors can be subcategorized into the following industries:

- **Consumer Discretionary**: Auto Components, Automobile, Household Durables, Leisure Products, Textiles Apparel and Luxury Goods, Hotels Restaurants and Leisure, Diversified Consumer Services, Media, Distributors, Internet and Direct Marketing Retail, Multiline Retail and Specialty Retail
- **Financials**: Banks, Thrifts and Mortgage Finance, Diversified Financial Services, Consumer Finance, Capital Markets, Insurance and Equity Real Estate Investment Trusts (REITs)
- **Information Technology**: Real Estate Management and Development, Internet Software and Services, IT Services, Software, Communications Equipment Technology Hardware Storage and Peripherals and Electronic Equipment Instruments and Components

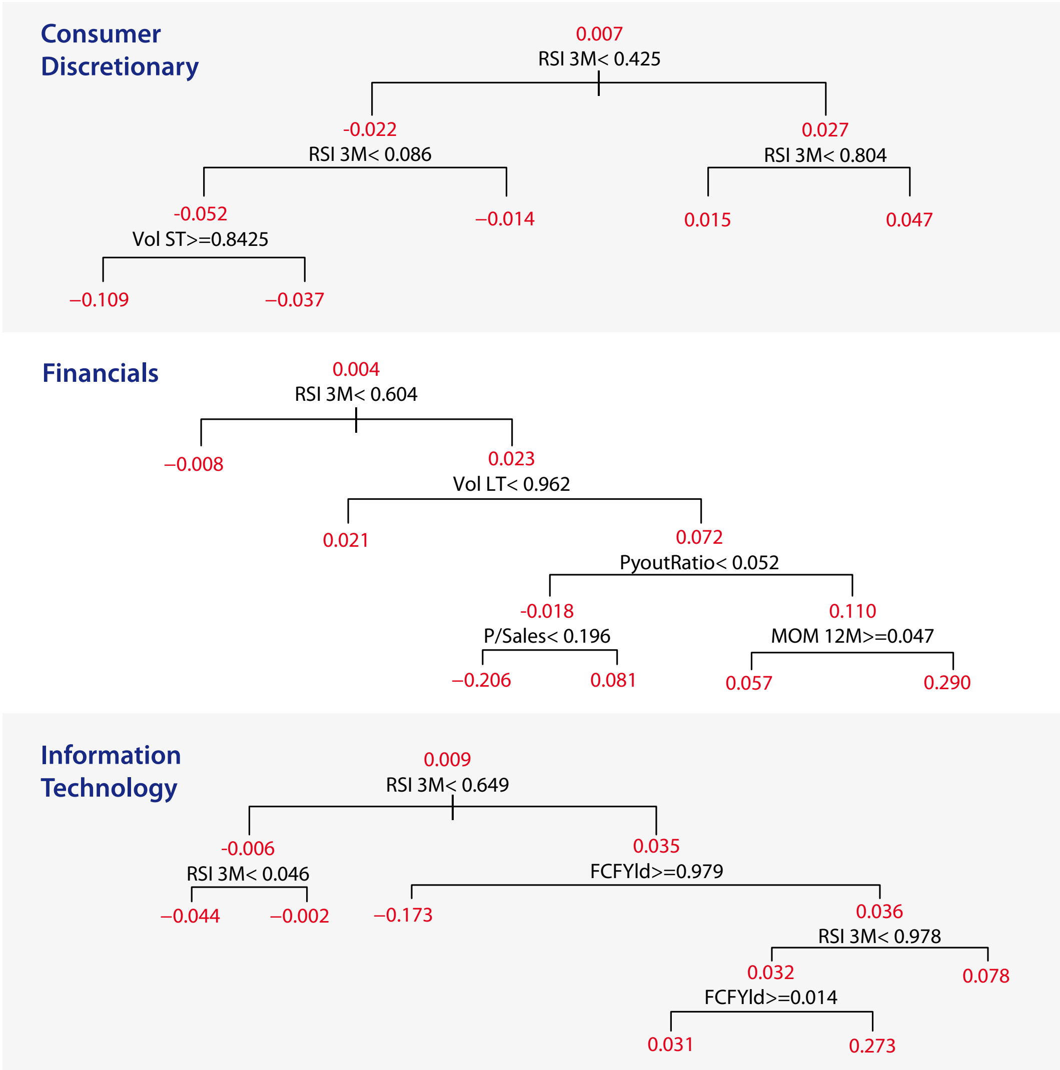

Figure 3: **Regression trees with 1 month future returns as dependent variable**. The splitting criteria are indicated in black and the average future 1 month returns are displayed in red for each node. The cluster corresponding to the criteria lies to the left of the node and the opposite cluster is located to its right. The leaves are ordered so that returns display an overall increasing pattern. The trees are pruned to improve readability.

4.3 Implication for portfolio choice

Given the strong evidence that the RSI characteristic is the major driver of future returns, we intend to quantify the economic gain that an investor can hope for when constructing investment policies based on this variable. The strategy is implemented as follows. At the beginning of each period[9], we rank the stocks according to the RSI 3M variable and we keep only the top $q\%$, where q spans the set $\{50, 40, 30, 20, 10\}$. For $q = 40$, this leads to portfolios of 182 stocks on average, which allows for a reasonable diversification. Then, each stock is weighted uniformly in the portfolio (the 1/N rule is known to be efficient, see e.g., DeMiguel et al. (2009)). At the end of the period, we compute the corresponding return and construct the new portfolio weights with the updated RSI 3M. To put our results in perspective with other selection schemes, we

[9]In order to assess the sensitivity of our method, we consider four different holding periods: 1, 3, 6 and 12 months.

compare the strategy based on the RSI 3M to that based on the classical momentum (computed between 12 and one month prior to portfolio formation).

In Table 2, we report the following five performance metrics:

- **Annualized arithmetic return**: $r = \frac{1}{T\delta_\Delta}\sum_{t=1}^{T} r_t(\Delta)$, where the returns depend on the holding period Δ (1, 3, 6 or 12 months) and δ_Δ is the annualizing constant,
- **Annualized volatility**: $\sigma = \sqrt{\frac{1}{T\delta_\Delta}\sum_{t=1}^{T}(r_t(\Delta) - r)^2}$,
- **Sharpe ratio**: r/σ,
- **Annualized turnover**: $\text{Turn} = \frac{1}{TN\delta_\Delta}\sum_{t=2}^{T}\sum_{n=1}^{N}|w_{n,t} - w_{n,t-1}|$, where $w_{n,t}$ is the weight in the portfolio of stock n at time t,
- **Annualized return, net of transaction costs**: $r^{TC} = r - 0.005 \times \text{Turn}$.[10]

	RSI 3M					MOM 12M					1/N
	q=50	q=40	q=30	q=20	q=10	q=50	q=40	q=30	q=20	q=10	
	1 month holding period										
r	0.276	0.320	0.368	0.436	0.529	0.082	0.083	0.087	0.092	0.112	0.121
σ	0.182	0.186	0.193	0.202	0.219	0.148	0.149	0.151	0.163	0.185	0.179
SR	1.516	1.718	1.906	2.154	2.414	0.559	0.556	0.577	0.562	0.605	0.677
Turn	11.194	13.369	15.586	17.797	20.421	3.732	4.433	5.239	6.162	7.174	0.071
r^{TC}	0.220	0.253	0.290	0.347	0.427	0.064	0.061	0.061	0.061	0.076	0.121
	3 month holding period										
r	0.120	0.133	0.149	0.171	0.204	0.057	0.059	0.062	0.064	0.073	0.087
σ	0.128	0.129	0.131	0.133	0.137	0.110	0.111	0.113	0.121	0.136	0.138
SR	0.938	1.032	1.142	1.282	1.485	0.521	0.526	0.546	0.526	0.534	0.629
Turn	3.734	4.460	5.197	5.938	6.808	1.248	1.482	1.750	2.062	2.397	0.023
r^{TC}	0.101	0.111	0.123	0.141	0.170	0.051	0.051	0.053	0.054	0.061	0.087
	6 month holding period										
r	0.081	0.090	0.098	0.110	0.129	0.049	0.051	0.052	0.053	0.059	0.075
σ	0.111	0.110	0.110	0.112	0.115	0.102	0.103	0.106	0.111	0.127	0.128
SR	0.734	0.813	0.886	0.981	1.128	0.484	0.491	0.495	0.477	0.468	0.587
Turn	1.867	2.230	2.597	2.969	3.403	0.624	0.740	0.873	1.032	1.194	0.013
r^{TC}	0.072	0.079	0.085	0.095	0.112	0.046	0.047	0.048	0.048	0.053	0.075
	12 month holding period										
r	0.069	0.073	0.077	0.084	0.094	0.050	0.051	0.053	0.055	0.061	0.075
σ	0.107	0.108	0.105	0.106	0.109	0.101	0.104	0.108	0.117	0.139	0.136
SR	0.639	0.683	0.736	0.790	0.866	0.491	0.492	0.492	0.472	0.436	0.553
Turn	0.933	1.115	1.299	1.484	1.700	0.314	0.372	0.439	0.520	0.599	0.008
r^{TC}	0.064	0.068	0.071	0.076	0.086	0.048	0.049	0.051	0.052	0.058	0.075

Table 2: **Performance indicators of portfolio strategies**.

Since we know that the RSI 3M is strongly discriminating, it is only fair that the small portfolios that concentrate on the top decile have the highest returns: loosely speaking, they are the "purest" in the RSI 3M characteristic. More precisely, the returns are strictly decreasing in q. We note that this is also true for the momentum portfolios which confirms that momentum is indeed a priced factor. In addition, for both types of policies, the volatility is also decreasing with q for the shortest holding periods (one and three months). For the longest holding periods, the volatility is rather

[10]Both the definition of turnover and the magnitude of transaction costs originate from Brandt et al. (2009).

stable across q, even though it peaks for $q = 10$. The net effect in terms of Sharpe ratio (SR) is positive for the RSI 3M-based portfolios: the SR decreases with q. With regard to momentum portfolios, it is hard to draw clear conclusions apart from the fact that the SR are rather homogenous across q.

Turning transaction costs (TC), the mechanical effect of choosing fewer stocks increases asset rotation, hence the turnover decreases with q. Portfolios associated with low q (i.e. fewer stocks) are thus more profitable, but the corresponding transaction costs are much higher. The returns net of TC, though, are still decreasing with q.

When comparing holding periods, we see that the results are smoothed for the longer periods. Returns are higher when portfolios are updates more frequently, which is known as the rebalancing effect.[11] In practice, investors seldom rebalance their portfolios every month, but rather every three or six month.

Finally, it is valuable to compare these metrics with the equally-weighted portfolio of all stocks, which is a suitable benchmark (e.g., DeMiguel et al. (2009) and Plyakha et al. (2016)). Through this reference point, the major difference between RSI 3M-based and momentum-based portfolios materializes: the RSI 3M portfolios are all more profitable than the benchmark for holding periods from one to six months. Apart for one exception, this is true even when transaction costs are taken into account.

In what follows, we provide supplementary results for a reasonable policy. We set the rebalancing frequency to once every quarter and we fix the proportion of retained stocks to $q = 40(\%)$. In Figure 4, we plot the compounded cumulative returns (i.e. portfolio values normalized to one at inception) when returns have been adjusted for transaction costs. We compare three strategies: the portfolio based on the RSI 3M, the portfolio based on 12-month momentum and the 1/N benchmark. Given the results of Table 2, the TC-adjusted of these strategies is, respectively, 11.1%, 5.1% and 8.7%. The RSI strategies outperforms the momentum one by 6% annually, and the 1/N strategy by 2.4%.

The first strategy (blue line) obviously has an impressive track-record. Nevertheless, it is apparent that its outperformance stems from the latest period (2008 onwards), which is coherent with the dynamics of relative importance shown in Figure 2.[12] In the early years of the sample, the equally-weighted portfolio (which is a proxy of the market portfolio) takes advantage of the post internet bubble economic boom.

Surprisingly, the (long-only) momentum strategy displays the worst results of all. The main reason for that lies probably in so-called momentum crashes (Barroso and Santa-Clara (2015) and Daniel and Moskowitz (2016)). This phenomenon relies on the following rationale. After a financial crash, the losers are the stocks with the highest upward potential, while for the winners, it is the opposite. Thus, in the aftermath of a strong market downturn, the momentum strategy places bets in the wrong way and experiences strong negative returns. This arises as a plausible explanation for the disappointing levels of the green curve in Figure 4. Indeed, in the early years of the sample (2002-2003), betting on winners is counterproductive because losers are more likely to soar after the internet bubble burst. The most salient materialization of the momentum crash effect comes in 2009-2010 after the 2008 crisis. While both the 1/N and the RSI-based strategies are able to take advantage of the recovery, it is not the case

[11] See for instance the theoretical analysis of Liu and Strong (2008) and the empirical work of Plyakha et al. (2016) in the case of value- and equally-weighted portfolios.

[12] Notwithstanding a big difference in the data sample, this result is coherent with the conclusions of Han et al. (2013) in which short-term moving averages are found to generate very profitable trading strategies.

for the momentum strategy. The plus '+' line on Figure 4 shows the ratio between the 1/N portfolio and the momentum portfolio. The phases of strong increase for this line correspond to the 2002-2003 and 2009-2010 post-crisis periods. On the other hand, the RSI 3M portfolio seems immunized against the momentum crash effect, but we cannot put forward a rational explanation for this finding.

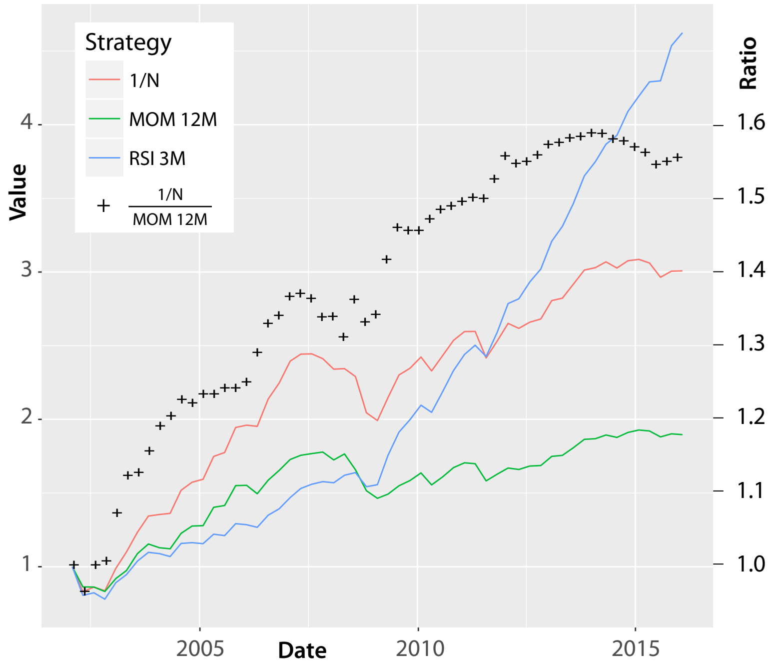

Figure 4: **Time-series of portfolio values**. In color, we plot the values of the portfolios; their scale is on the left. They are rebalanced quarterly and correspond to the case $q = 40$ in Table 2. Transaction costs are subtracted to returns. The dotted "+" curve (with scale on the right) shows the ratio between the value of the 1/N portfolio and the portfolio based on the 12 month momentum.

5 Conclusion

We perform regression trees to establish which firm characteristics are likely to drive future returns. One technical indicator, the Relative Strength Index, computed over three month, emerges as the primary splitting variable. This is true for one, six and twelve month forward returns and also when firms are grouped according to their industry focus. When looking at the relative importance of variables through time, we find that the three-month RSI has gradually gained significance after 2007. This result suggests that investors should pay attention to this variable when building portfolio strategies. We propose one such strategy and a reasonable choice of parameters leads to an annual gain of 2.4% over the equally-weighted portfolio. Consequently, we believe future research should further investigate both other possible nonlinear effects of firm characteristics over future returns and also international markets.

A Data

K	Short Name	Description
		Fundamental / Accounting
1	BuyBackYld	Buyback Yield
2	Cash Conversion	Cash to EBIT ratio
3	Debt/Equity	Debt on Equity
4	DivYld	Dividend Yield
5	EV/EBITDA	Enterprise Value to EBITDA ratio
6	FCFYld	Free Cash Flow Yield using FCF per Share
7	GrProfitGrowth	Gross Profits Growth (12 months)
8	Net Debt	Net Debt
9	NetDbtYld	Net Debt Yield
10	Net Prt Margin	Net Profit Margin
11	Op Prt Margin	Operating Profit Margin
12	P/B	Price to Book ratio
13	P/E	Price-Earnings ratio
14	P/Sales	Price to Sales ratio
15	PyoutRatio	Payout Ratio (Divided Per Share / Earnings Per Share)
16	ROC	Return On Capital
17	ROE	Return On Equity
18	ShareHldrYld	Total Shareholer Yield
		Momentum
19	MOM6	Momentum 6-1 months
20	MOM12	Momentum 12-1 months
21	RSI 3	RSI over 3 months
22	RSI 12	Relative Strength Index over 12 months using weekly returns
		Volatility
23	Vol LT	Ex-post volatility over 156 weeks
24	Vol ST	Ex-post volatility over 52 weeks
		Liquidity
25	ADV20	Average Daily Volume over 20 days
26	ADV60	Average Daily Volume over 60 days
27	ADV12M	Average Daily Volume over 12 months
28	BidAsk	Bid-Ask Spread
29	MKTCAP	Market Capitalization
30	Share Turn	Share turnover ratio

Table 3: Summary of firm characteristics

Characteristic	Minimum	Median	Maximum	Mean	Std. Dev.
ADV20	930.192	1.063e8	2.492e9	1.890e8	2.532e8
ADV60	930.192	1.075e8	2.469e9	1.894e8	2.511e8
ADV12M	1367.234	1.081e8	2.480e9	1.909e8	2.536e8
BidAsk	-0.018	0.000	0.021	0.001	0.002
BuyBackYld	-0.421	0.008	0.450	0.016	0.049
Cash Conversion	-32.678	1.002	31.480	1.102	2.003
Debt/Equity	-8067.549	54.772	7452.023	97.387	402.237
DivYld	0.000	0.019	0.186	0.023	0.019
EV/EBITDA	-486.896	10.277	506.402	11.682	22.591
FCFYld	-1.187	0.049	1.287	0.050	0.111
GrProfitGrowth	-10.133	0.067	10.220	0.093	0.454
MKTCAP	5.50e-05	1.080e4	3.735e5	2.428e4	4.054e4
MOM12	-0.995	0.073	2.764	0.086	0.332
MOM6	-0.987	0.040	1.711	0.040	0.215
NetDebt	-2.173e4	1781.000	2.158e4	4240.242	18764.315
NetDbtYld	-1.928	-0.001	1.894	-0.013	0.161
Net Prt Margin	-6.559	0.093	4.345	0.088	0.309
Op Prt Margin	-5.986	0.165	5.615	0.177	0.254
P/B	-218.456	2.623	227.075	3.560	11.572
P/E	-2707.395	18.028	2643.030	20.580	106.513
P/Sales	-48.988	1.937	68.337	2.925	3.637
PyoutRatio	-16.511	0.307	17.463	0.394	1.070
ROC	-1.264	0.136	1.594	0.155	0.139
ROE	-19.913	0.140	23.333	0.152	0.734
RSI 3	3.977	51.905	98.214	51.885	10.235
RSI 12	27.133	51.659	95.122	51.627	5.865
ShareHldrYld	-0.690	0.010	0.706	0.007	0.061
Share Turn	0.000	0.007	0.092	0.010	0.009
Vol LT	0.105	0.292	1.591	0.330	0.158
Vol ST	0.018	0.267	1.683	0.310	0.169

Table 4: **Descriptive statistics of firm characteristics.**

References

Ammann, M., Coqueret, G., Schade, J.-P., 2016. Characteristics-based portfolio choice with leverage constraints. Journal of Banking & Finance 70, 23–37.

Asness, C. S., Moskowitz, T. J., Pedersen, L. H., 2013. Value and momentum everywhere. Journal of Finance 68 (3), 929–985.

Atkinson, E. J., Therneau, T. M., 2015. An introduction to recursive partitioning using the rpart routines. Rochester: Mayo Foundation.

Bansal, R., Viswanathan, S., 1993. No arbitrage and arbitrage pricing: A new approach. Journal of Finance 48 (4), 1231–1262.

Banz, R. W., 1981. The relationship between return and market value of common stocks. Journal of Financial Economics 9 (1), 3–18.

Barroso, P., Santa-Clara, P., 2015. Momentum has its moments. Journal of Financial Economics 116 (1), 111–120.

Brandt, M. W., Santa-Clara, P., Valkanov, R., 2009. Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. Review of Financial Studies 22 (9), 3411–3447.

Daniel, K., Moskowitz, T. J., 2016. Momentum crashes. Journal of Financial Economics 122 (2), 221–247.

Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross sectional variation in stock returns. Journal of Finance 52 (1), 1–33.

Daniel, K., Titman, S., Wei, K., 2001. Explaining the cross-section of stock returns in Japan: Factors or characteristics? Journal of Finance 56 (2), 743–766.

DeMiguel, V., Garlappi, L., Uppal, R., 2009. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? Review of Financial Studies 22 (5), 1915–1953.

DeMiguel, V., Martin-Utrera, A., Nogales, F. J., Uppal, R., 2017. A portfolio perspective on the multitude of firm characteristics. SSRN Working Paper.

Fama, E. F., French, K. R., 1992. The cross-section of expected stock returns. Journal of Finance 47 (2), 427–465.

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.

Friedman, J., Hastie, T., Tibshirani, R., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition. Springer.

Goyal, A., 2012. Empirical cross-sectional asset pricing: A survey. Financial Markets and Portfolio Management 26 (1), 3–38.

Graham, B., Dodd, D. L., 1934. Security analysis: Principles and technique. McGraw-Hill.

Green, J., Hand, J. R., Zhang, F., 2017. The characteristics that provide independent information about average US monthly stock returns. Review of Financial Studies Forthcoming, X–X.

Green, J., Hand, J. R., Zhang, X. F., 2013. The supraview of return predictive signals. Review of Accounting Studies 18 (3), 692–730.

Han, Y., Yang, K., Zhou, G., 2013. A new anomaly: The cross-sectional profitability of technical analysis. Journal of Financial and Quantitative Analysis 48 (05), 1433–1461.

Han, Y., Zhou, G., Zhu, Y., 2016. A trend factor: Any economic gains from using information over investment horizons? Journal of Financial Economics 122 (2), 352–375.

Harvey, C. R., Liu, Y., Zhu, H., 2015. ... and the cross-section of expected returns. Review of Financial Studies 29 (1), 5–68.

Hjalmarsson, E., Manchev, P., 2012. Characteristic-based mean-variance portfolio choice. Journal of Banking & Finance 36 (5), 1392–1401.

Jacobs, H., 2015. What explains the dynamics of 100 anomalies? Journal of Banking & Finance 57, 65–85.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. Journal of Finance 48 (1), 65–91.

Jegadeesh, N., Titman, S., 2001. Profitability of momentum strategies: An evaluation of alternative explanations. Journal of Finance 56 (2), 699–720.

Kara, Y., Boyacioglu, M. A., Baykan, Ö. K., 2011. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange. Expert Systems with Applications 38 (5), 5311–5319.

Kim, S.-H., Kim, D., 2014. Investor sentiment from internet message postings and the predictability of stock returns. Journal of Economic Behavior & Organization 107, 708–729.

Kocadagli, O., 2013. A novel nonlinear programming approach for estimating capm beta of an asset using fuzzy regression. Expert Systems with Applications 40 (3), 858–865.

Krauss, C., Do, X. A., Huck, N., 2016. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. European Journal of Operational Research.

Liu, W., Strong, N., 2008. Biases in decomposing holding-period portfolio returns. Review of Financial Studies 21 (5), 2243–2274.

Lu, T.-H., Chen, Y.-C., Hsu, Y.-C., 2015. Trend definition or holding strategy: What determines the profitability of candlestick charting? Journal of Banking & Finance 61, 172–183.

McLean, R. D., Pontiff, J., 2016. Does academic research destroy stock return predictability? Journal of Finance 71 (1), 5–32.

Menkhoff, L., 2011. Are momentum traders different? Implications for the momentum puzzle. Applied Economics 43 (29), 4415–4430.

Mingers, J., 1989. An empirical comparison of pruning methods for decision tree induction. Machine Learning 4 (2), 227–243.

Morgan, J. N., Sonquist, J. A., 1963. Problems in the analysis of survey data, and a proposal. Journal of the American Statistical Association 58 (302), 415–434.

Moritz, B., Zimmermann, T., 2016. Tree-based conditional portfolio sorts: The relation between past and future stock returns. Working Paper.

Neely, C. J., Rapach, D. E., Tu, J., Zhou, G., 2014. Forecasting the equity risk premium: the role of technical indicators. Management Science 60 (7), 1772–1791.

Peltomäki, J., Äijö, J., 2015. Cross-sectional anomalies and volatility risk in different economic and market cycles. Finance Research Letters 12, 17–22.

Plyakha, Y., Uppal, R., Vilkov, G., 2016. Equal or value weighting? Implications for asset-pricing tests. SSRN Working Paper.

Roll, R., 1977. A critique of the asset pricing theory's tests part I: On past and potential testability of the theory. Journal of Financial Economics 4 (2), 129–176.

Sorensen, E. H., Miller, K. L., Ooi, C. K., 2000. The decision tree approach to stock selection. Journal of Portfolio Management 27 (1), 42–52.

Subrahmanyam, A., 2010. The cross-section of expected stock returns: What have we learnt from the past twenty-five years of research? European Financial Management 16 (1), 27–42.

Suh, S., Song, W., Lee, B.-S., 2014. A new method for forming asset pricing factors from firm characteristics. Applied Economics 46 (28), 3463–3482.

Taylor, N., 2014. The rise and fall of technical trading rule success. Journal of Banking & Finance 40, 286–302.