Fundamental Signals Strategy

Daniel Cohn, Chase Navellier, Thomas Rogers

MS&E 448 - June 2018

1 Abstract

Our project explores the predictive power of "quality" fundamental signals on equity performance. We first analyzed individual fundamentals' predictive power by performing linear regressions of 40 fundamentals against stock price changes for the following three and six-month periods using over 6,500 U.S. equities from 1996-2011. We then performed an elastic net regression on the 12 fundamentals that indicated greatest predictive potential in order to determine the optimal weighting of these fundamentals to construct a fundamental signal-based trading strategy. Lastly, we tested our algorithm within Quantopian using data from 2013 to the present, producing an annualized Sharpe ratio of 0.76 and a 14% greater return than that of the S&P 500 over the same period.

2 Introduction

Fundamental stock ratios, which refer to any quantitative metrics about a company and its stock price that are agnostic of company size, have been used to determine stock valuations and predict future price movements for as long as equities have been publicly traded. According to the classical academic theories of Markowitz' modern portfolio theory (MPT) and the capital asset pricing model (CAPM), greater returns from an investment portfolio are necessarily accompanied by greater risk, manifesting as greater expected volatility and greater market exposure ("beta") respectively. [1] Of course, some companies are superior to others from performance and safety standpoints (e.g. companies that generate relatively greater profits and have relatively lower levels of debt). Since higher equity valuations translate to lower expected returns, according to academic theory the stock prices of fundamentally superior companies (hence-force referred to as "quality" companies) should possess enough of a "premium" that the greater expected profit and lower risk are offset such that all equities return the same amount for a given level of risk. In CAPM, stocks that require an intercept ("alpha") in their regression against market exposure in order to explain returns are considered to have beaten or lost to the market, meaning that statistically significant alphas should not be possible for even the most quality of companies.

Surprisingly, however, it has been repeatedly noted both in practice and through academic literature that equity portfolios constructed by overweighting quality stocks (of various definitions) have historically generated statistically significant alpha. Thus, the ideas that (1) identification of quality companies can lead to market-beating risk-adjusted performance and that (2) this phenomenon can be captured largely or entirely quantitatively served as the underlying motivations for our project this quarter. Specifically, we based our project off of two papers that claimed to demonstrate market outperformance using varying methodologies of quality signal definition and portfolio construction.

	(1) Sharpe Ratio	(2) β	(3) β^-	(4) *Skewness*	(5) Proba $(r_t < -2\sigma)$	(6) Signal Persistence
Market - short rate	.47	1	1	-.13	.031	.
Low vol	.43	-.015	0	-.06	.032	.99
Book to Market	.2	.029	.11	.035	.025	.98
Repurchasers	.55	.01	.04	-.053	.019	.96
Momentum	.43	-.041	-.1	-.007	.025	.88
Industry Leaders	.48	-.016	-.14	.008	.029	.15
Accruals	.77	.014	-.027	.027	.018	.95
ROE	.55	-.025	-.033	.021	.01	.97
Cash-Flows	1.2	-.016	-.055	.06	.021	.97
ROA	.46	-.025	-.054	.08	.01	.99

Figure 1: Performance of Various Trading Signals

In *The Excess Returns of "Quality" Stocks: A Behavioral Anomaly* by Bouchaud, Cilberti, Landier, Simon, and Thesmar (2016), the authors establish the signal persistence of fundamental metrics, defined as the regression coefficient of a fundamental value against that of the previous month [Fig. 1, Column 6]. [2] The notably high persistence in most of the fundamental signals considered makes intuitive sense given the relatively static nature of organizations and the markets and governments in which they operate, and confirms the potential utility of fundamental metrics when valuing the future cash flows of their underlying companies. The authors also declare that companies whose quality is defined by operating cash flow over total assets widely outperformed the historical market. Using the largest 1,500 US stocks from 1990-2012 as a universe, they demonstrate that a market-neutral long/short portfolio whose weightings are proportional to relative quality value would have achieved a statistically significant Sharpe ratio (a volatility-adjusted return metric) of 1.17 [Fig. 2]. This is considerably above the Sharpe ratio of 0.47 of the overall market during this period, though it should be noted that no transaction or borrow costs were factored in.

Having established that "the 'quality anomaly' is one of the capital markets' strongest reported anomalies and has a long tradition among investors," the authors then aim to understand why this anomaly persists despite its long

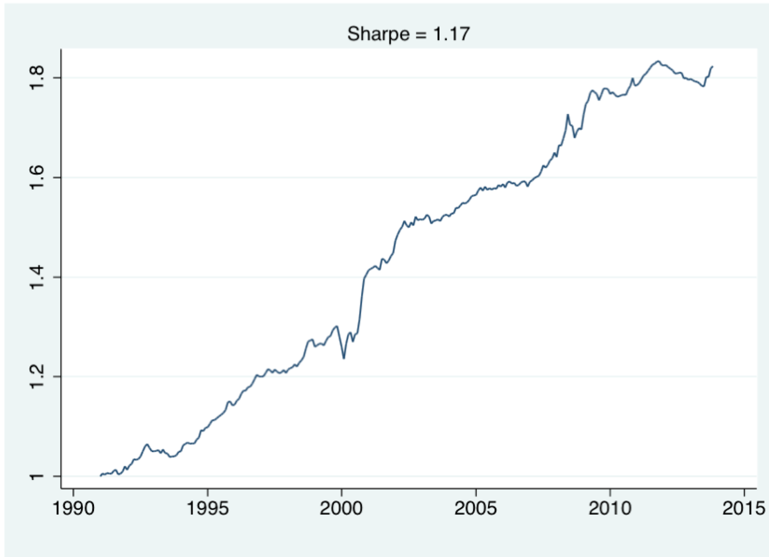

Figure 2: Cumulative Returns Using OCF/Assets Signal

history and relatively wide acceptance. They consider the possibility that the excess returns of quality companies are not actually excessive when viewed on a risk-adjusted basis, as well as the possibility that the returns are indeed an anomaly that is best explained through the lens of the behavioral psychology that underpins human markets. As evidenced by the low betas of various fundamental quality signals in Figure 1, the authors ultimately find no evidence of quality companies being correlated with higher market risk; indeed, a propensity for the market to "flee to quality" during times of distress even results in a lower probability of significant downturn [Fig. 1, Columns 2 & 5]. With the risk explanation discredited, the rest of the paper argues that the quality anomaly is largely explained by a systematic focus on the wrong metrics and a psychological tendency to update opinions more slowly than Bayes' rule would suggest, leading to consistent overvaluation of "junk" and undervaluation of "quality."

In *Quality Minus Junk* by Asness, Frazzini, and Pedersen (2013), the authors consider a "quality" firm as any publicly traded company that is "safe, profitable, growing, and well-managed." [3] The authors define four categories of quality, profitability, growth, safety, and payout, each constructed through equal weighting of relevant z-scored fundamentals. Profitability was considered by looking at each stock's average rank across "gross profits, margins, earnings, accruals, [and] cash flows." Growth was defined as the 5-year change amongst profitability metrics. Safety was considered by looking at stocks with "low leverage, low volatility of profitability, and low credit risk." Lastly, payout was categorized as "fraction of profits paid out to shareholders." Using CAPM as well as more complex pricing models that include company size, valuation, and momentum factors, portfolios constructed from each category signal as well as a z-scored, equally-weighted combination of the four signals ("Quality Minus Junk" or QMJ) consistently found statistically significant alpha by going long the top 30% and short the bottom 30% of all actively traded U.S. equities from 1956-2012. This phenomenon was also present in a shorter sample encompassing the global universe of equities [Fig. 3].

	Panel A: Long Sample (U.S. , 1956 - 2012)					Panel B: Broad Sample (Global , 1986 - 2012)				
	QMJ	Profitability	Safety	Growth	Payout	QMJ	Profitability	Safety	Growth	Payout
Excess Returns	**0.40**	**0.27**	**0.23**	0.12	**0.31**	**0.38**	**0.34**	0.19	0.02	**0.38**
	(4.38)	(3.81)	(2.06)	(1.63)	(3.37)	(3.22)	(3.30)	(1.33)	(0.24)	(3.41)
CAPM-alpha	**0.55**	**0.33**	**0.42**	0.08	**0.46**	**0.52**	**0.43**	**0.34**	0.02	**0.49**
	(7.27)	(4.78)	(4.76)	(1.06)	(6.10)	(5.75)	(4.61)	(3.07)	(0.18)	(5.29)
3-factor alpha	**0.68**	**0.45**	**0.59**	**0.20**	**0.43**	**0.61**	**0.53**	**0.50**	0.14	**0.44**
	(11.10)	(7.82)	(8.68)	(3.32)	(6.86)	(7.68)	(6.11)	(5.40)	(1.92)	(5.17)
4-factor alpha	**0.66**	**0.53**	**0.57**	**0.38**	**0.21**	**0.45**	**0.49**	**0.39**	**0.29**	**0.19**
	(10.20)	(8.71)	(7.97)	(6.13)	(3.43)	(5.50)	(5.34)	(4.00)	(3.91)	(2.26)
MKT	**-0.25**	**-0.11**	**-0.34**	**0.05**	**-0.20**	**-0.24**	**-0.16**	**-0.28**	0.00	**-0.18**
	(-17.02)	(-8.08)	(-20.77)	(3.35)	(-14.47)	(-14.36)	(-8.33)	(-13.74)	(-0.06)	(-10.50)
SMB	**-0.38**	**-0.21**	**-0.41**	**-0.05**	**-0.30**	**-0.33**	**-0.20**	**-0.31**	**-0.18**	**-0.23**
	(-17.50)	(-10.21)	(-17.00)	(-2.53)	(-14.82)	(-9.46)	(-5.07)	(-7.48)	(-5.62)	(-6.58)
HML	**-0.12**	**-0.28**	**-0.23**	**-0.44**	**0.39**	-0.01	**-0.16**	**-0.22**	**-0.38**	**0.36**
	(-5.03)	(-12.16)	(-8.50)	(-18.81)	(16.68)	(-0.31)	(-3.95)	(-5.23)	(-11.62)	(9.89)
UMD	0.02	**-0.07**	0.01	**-0.17**	**0.21**	**0.15**	0.03	**0.10**	**-0.14**	**0.24**
	(0.82)	(-3.80)	(0.64)	(-8.55)	(10.79)	(5.54)	(1.01)	(3.07)	(-5.64)	(8.57)
Sharpe Ratio	0.58	0.51	0.27	0.22	0.45	0.62	0.63	0.26	0.05	0.66
Information Ratio	1.46	1.25	1.14	0.88	0.49	1.16	1.13	0.84	0.83	0.48
Adjusted R2	0.57	0.37	0.63	0.40	0.60	0.60	0.34	0.58	0.35	0.52

Figure 3: Market Outperformance of Various Quality Categories

Given that both of these studies present some level of market-beating statistical significance amongst various quality signals, we were encouraged to conduct our own research into fundamental metrics. We noted that the first paper only considered individual fundamental signals in isolation, and that the second paper evenly weighted each signal in each category, as well as evenly weighted each category when producing an overall QMJ signal. Furthermore, the decisions as to which fundamentals to include were chosen discretionarily from a wide array of previous literature. Therefore, we saw the opportunity to uncover an improved quality signal by (1) researching the predictive power of several fundamentals from the same environment statistical environment and (2) determining the fundamental weightings of our signal in an optimal manner.

An additional avenue we believed to have the potential to improve the predictive power of our overall signal was the incorporation of "value" fundamentals that relate details of a company's operation to its stock price. This concept of a "quality at a reasonable price" (QARP) metric is supported in *Quality Minus Junk* by the fact that the historic outperformance of quality is inversely related with high correlation to the quality premium that was paid to own said quality securities. That is to say that, quite intuitively, the periods when quality companies are at their cheapest pose the greatest potential for outsized market returns using this strategy. Accordingly, incorporation of valuation metrics into the overall fundamental signal should provide additional ammunition in finding companies that are relatively undervalued. Quality Minus Junk reported the results of a basic QARP signal constructed as a linear combination of the QMJ factor and the price-to-book value. However, they never considered valuation metrics beyond price-to-book, and they found the specific linear combination by optimizing for the Sharpe ratio over their test set, which we considered statistically objectionable. As a result, we also believed there to be significant room for improvement in the overall fundamental signal through incorporation of "value" fundamentals. Regardless, their result of a QARP Sharpe ratio of 0.7 using the US universe of stocks served as a helpful point of reference.

3 Data

We chose to remain within the universe of all tradable US stocks in order to maximize the number of data points in our analysis while conditioning on a single regulatory and reporting environment and staying consistent with the prior literature. Initially, our group hoped to use the data accessible through Quantopian to build and train our model so that we could easily test with the same tools provided and ensure data reporting consistency. However, the datasets available through Quantopian were sparse and not mutable enough for our project. Additionally, the rigid infrastructure prevented facile research. Instead, we relied on two datasets from Wharton Research Data Services. The first dataset contained over 70 different monthly fundamental ratios for various firms from 1996 to 2011. The second dataset contained information on firms' monthly stock price and returns over the same period. When combined, these datasets allowed us to examine the relationship between a firm's fundamentals and its future return.

3.1 Data Cleaning and Preparation

Before we could begin our analysis, there were several issues with the dataset that had to be addressed. First and foremost, our group needed to merge the two datasets. With over 1.4 million rows and seventy columns in each dataset, this proved a nontrivial task. In order to properly merge the two datasets, we aligned the rows by firm and date. Each firm was identified by a PERMNO, or permanent number. This field was particularly useful in that it stayed constant indefinitely, in spite of ticker changes or delistings. The two datasets each contained information on over 10,000 firms; however, they only had roughly 6,500 firms in common. As a result, our master dataset contained information on these 6,500 firms reported monthly over the 16-year period from 1996 to 2011 (over 650,000 rows in total).

Once all the necessary data was combined, we focused on cleaning the master dataset. Specifically, there were two issues we needed to handle: missing data and extreme values. Because of the wide ranges of possible values and potential significance of even a single predictor, we decided to simply remove rows with missing data instead of substituting mean or expected values. Additionally, several of the data points were several orders of magnitude greater than the mean, resulting in impractical values. To handle these points and prevent them from over-leveraging our future models, we removed all rows containing data points that were more than five standard deviations away from the field's mean. Once we accounted for missing data and extreme values, the master dataset contained just over 425,000 rows.

After restricting the dataset to full rank rows with values in a reasonable range, we created two new fields that would serve as our dependent variables in the model building process, namely future returns over a three and six-month window. For example, if a row's fundamental data came from the month of January 2001, the three-month return field would contain the compounded returns from February 2001 to April 2001. We conservatively chose to start the return window during the fundamentals' following month since we were not able to verify that the pricing and fundamentals information during the same month were recorded on the same date and therefore wanted to prevent any potential lookahead bias. Overall, the final dataset contained the company PERMNO, the reporting date of the fundamental data, the fundamental ratios, and the future returns.

3.2 Data Transformations

The eventual goal of our data analysis was to combine the predictive power of our individual fundamental signals into a single signal capable of predicting future returns with maximal accuracy. In order to create a single signal, the rest of our predictors had to be in a form suitable for combination where their differing units would not cause issues. To achieve this unit-less standardization, at each time period we individually z-scored all of the fields in our dataset. Following this step, our dataset was fully cleaned and treated.

4 Individual Fundamental Analysis

4.1 Identifying Significant Fundamentals

Among the seventy fundamentals we had access to, we selected forty that had some potential to indicate value or profitability/safety quality in order to avoid data snooping. Because our return data was limited to only price changes, we lacked dividend information and therefore did not include fundamentals indicative of payout quality. We also did not include growth fundamentals since they had not shown statistically significant CAPM alpha (the alpha ultimately calculated within the Quantopian backtester) and since we didn't believe our dataset spanned enough history to confidently make conclusions about five-year growth of profitability fundamentals.

Because we had z-scored the returns and fundamentals at each time point, we were able to combine the data from multiple time periods into a single regression by negating the effect of market conditions on fundamental and return values. We determined that a simple linear regression of each fundamental against future returns would provide the relative predictive power we hoped to determine without overfitting the data. While we had little reason to believe that the underlying function was indeed linear, we decided that a linear regression made the most sense given the varied and opaque complexities of the relationships governing each fundamental's relationship to return. We also generally confirmed that the distribution of returns across the fundamentals was approximately Gaussian with similar variance (although a more rigorous adherence to this assumption may improve the results moving forward). [4] We then performed individual regressions on each of the forty fundamentals against both three and six-month returns (eighty regressions in total) in order to gauge the time over which stock prices might take to react to a signal. Our program automatically removed any regression whose coefficient was below the 5% significance level and then ranked the fundamentals according to the magnitude of their coefficient [Fig. 4]. We were pleased to see that the top predictive fundamentals for both the three and six-month windows were largely the same, which indicated both strong signal persistence and signal relevance over time. We were also pleased to see almost all of the coefficients' signs were aligned with whether the fundamental was expected to positively or negatively correlate with returns; this served as an important reality-check going forward. The top twelve fundamentals to appear in both rankings were then selected as the basket with which to construct our overall fundamental signal.

4.2 Identifying Collinearity

After performing these individual regressions, we also looked into any relationships existing between our twelve predictors since many of our fundamental ratios share numerators or denominators, such as return on total assets and return on invested capital. Indeed, we suspected that there would be a high degree of collinearity between some predictors. To examine this relationship, we plotted the correlation between all twelve predictors [Fig. 5]. While the plot does reveal that there is indeed significant correlation between some of predictors, we did not remove any of them at this time. Because it wasn't clear which ones exactly to remove, we opted instead to combine the fundamentals via a model that would account for collinearity.

5 Signal Construction

5.1 Model Selection

There were several different models our group considered using to develop our strategy, ranging from complex neural networks to simple linear regressions. There were three key factors that we considered in deciding on a final model. First, due to the high-dimensionality of our training dataset the model would have to be resistant to overfitting. Second, because of the nature of our data the model must be trainable on predictors that are collinear. This immediately ruled out multiple linear regression since collinearity can make predictor coef-

3-Month Returns

```
----- roa -----
Coefficient:
 0.06107589945194121
----- ocf_lct -----
Coefficient:
 0.06040861674481992
----- cash_debt -----
Coefficient:
 0.04891110088785186
----- roce -----
Coefficient:
 0.04383086461667993
----- ps -----
Coefficient:
 -0.04321043338331969
----- roe -----
Coefficient:
 0.04104724503176234
----- bm -----
Coefficient:
 0.03521852212373669
----- GProf -----
Coefficient:
 0.029139526852746143
----- cash_lt -----
Coefficient:
 -0.023919730971725443
----- rect_act -----
Coefficient:
 0.0224087009511717
----- cfm -----
Coefficient:
 0.0209386574175141
----- npm -----
Coefficient:
 0.020396839450069388
----- quick_ratio -----
Coefficient:
 -0.01968701599174346
----- opmad -----
Coefficient:
 0.019303275543765385
----- opmbd -----
Coefficient:
 0.019157447726316716
----- pretret_noa -----
Coefficient:
 0.01866951511664456
```

6-Month Returns

```
----- ocf_lct -----
Coefficient:
 0.07518700043081517
----- roa -----
Coefficient:
 0.07167661067100652
----- cash_debt -----
Coefficient:
 0.059346288319372074
----- bm -----
Coefficient:
 0.05630492509935517
----- ps -----
Coefficient:
 -0.05553319230373514
----- roce -----
Coefficient:
 0.04704847747895677
----- roe -----
Coefficient:
 0.042899419904306
----- cash_lt -----
Coefficient:
 -0.03603221872125644
----- quick_ratio -----
Coefficient:
 -0.031061880720905436
----- GProf -----
Coefficient:
 0.03100606874826707
----- rect_act -----
Coefficient:
 0.03063337770423184
----- debt_assets -----
Coefficient:
 0.028155323934686654
----- curr_ratio -----
Coefficient:
 -0.026604203543925738
----- pretret_noa -----
Coefficient:
 0.0264272227406736
----- npm -----
Coefficient:
 0.026274752959098743
----- cfm -----
Coefficient:
 0.026150102935343908
```

Figure 4: Top Performing Individual Fundamentals

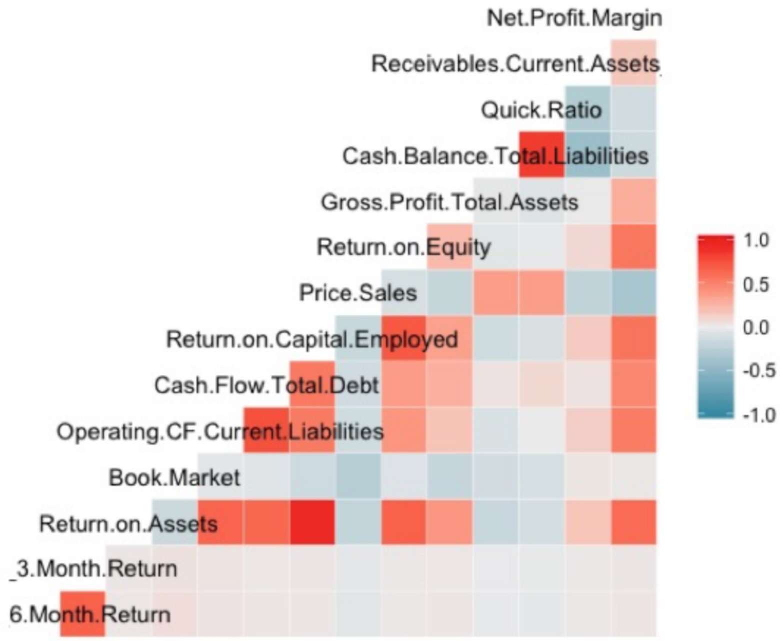

Figure 5: Collinearity Matrix of Top Fundamentals

ficients unstable. Lastly, the model had to be readily interpretable for both implementation and "sanity check" purposes, meaning more complex models would not suit our needs. Consequently, an interpretable, linear model would allow us to compare to prior literature and understand if the model was likely making mistakes.

In line with these requirements, our group decided to base our strategy around an elastic net model. An elastic net is a combination of the more commonly known LASSO and RIDGE regressions, both of which are themselves slightly modified multiple linear regressions. The elastic net model adds a penalty based on the number of predictors included, which in turn can bring predictor coefficients to zero. This reduces the variance of the model at the expense of model bias, helping increase resistance to overfitting. Additionally, in theory the model works by removing the coefficients that add the least value. Thus, coefficients whose information is contained within other predictors should be decreased or removed entirely by the model. Because of this, the elastic net model should also remove most collinearity, thereby making the model coefficients stable. Lastly, the model is highly interpretable as it returns an array of predictor coefficients. This is an extremely valuable feature of our model as it provides direct insight into the magnitude and direction of each predictor's effect on future returns.[5]

5.2 Predictor Selection

To view how prior predictor selection would impact the model performance, we trained models over two sets of predictors: the previously selected twelve fundamentals based on their individual performances and the full set of forty. We trained two models over both of these predictor sets using the two dependent variables of three and six-month returns, resulting in four models total. Training over the different predictor sets allowed us to generate optimal weightings of the previously selected top fundamentals while also giving the model an opportunity to prune predictors on its own out of the entire feature set. Additionally, running each regression against the two different dependent variables allowed us to test the length over which predictors would likely influence returns.

5.3 Creating Training and Validation Sets

Before training the four elastic net models, we first had to create training, validation, and test sets. Instead of creating a holdout set out of testing data, out-of-sample backtesting in Quantopian would serve to indicate relative model test performance. To create our training and validation sets, our group used 5-fold cross validation. The data was parsed into 5 different, roughly equal-length periods: 1996-1999, 2000-2002, 2003-2005, 2006-2008, and 2009-2011. While increasing the number of folds would have given us more data to train each iteration over, it would have also increase the likelihood of overfitting the model. Additionally, our group wanted ensure that each time period was at least three years so that we could specifically use the middle year of the period as our validation set for that fold. The reasoning behind this was that each data point contained either three months or six months of future return data, so if we had tested our model on months adjacent to our training data we would have introduced lookahead bias.

5.4 Model Tuning

Using this cross validation we then tuned the model hyperparameters. Specifically, these features were the complexity penalty, known as alpha, and the ratio of LASSO to RIDGE regression, known as the L1 Ratio. Both of these values can range from 0 to 1. An Alpha of 1 will result in all zero coefficients while an Alpha of 0 will result in a standard multiple linear regression. An L1 Ratio of 0 results in a typical LASSO Regression while an L1 Ratio of 1 results in a standard RIDGE Regression. In order to tune these parameters, each of the four models was trained using every combination of Alpha and L1 Ratio from 0 to 1, in increments of 0.01, for a total of 10,000 different possible combinations. To find the optimal combination, we found the pair that resulted in the lowest average MSE across the 5-folds. The below graphs show the change in MSE over the range of Alphas and L1 Ratios (when plotting Alphas, the optimal L1 Ratio is used and vice versa) [Fig. 6]. Considering the 10,000 different combinations of Alpha and L1 Ratio, the 5-fold cross validation, and the four different models we tuned, we eventually ran a total of 200,000 Elastic Net Regressions in search of the optimal models.

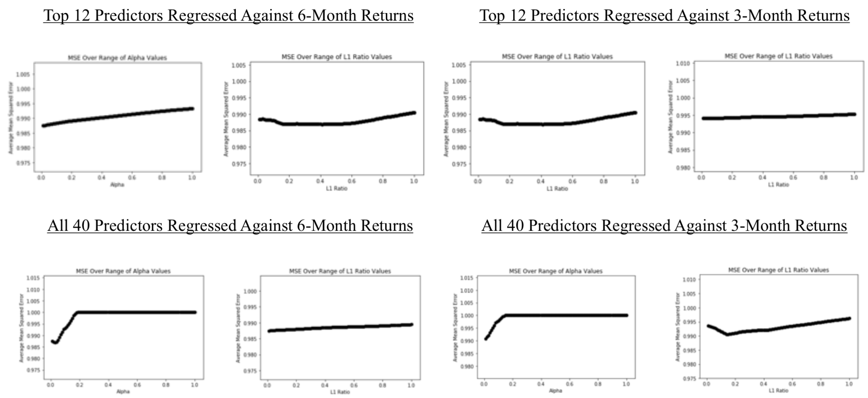

Figure 6: Elastic Net Hyperparameter Optimization

5.5 Elastic Net Results

For both models trained over the top fundamentals, the optimal parameter sets were an Alpha of 0.01 and L1 Ratio of 0.01. The minimum average MSE for the 6-month and 3-month models were 0.98746 and 0.94409 respectively (standard deviations). Based on these results, it appeared that the value of good fundamentals was immediate and had more effect over the next 3-months of returns than over the next 6-months. For the model trained over the full feature set against 6-month returns, the optimal parameters were an Alpha of 0.03 and an L1 Ratio of 0.41. The minimum average MSE achieved with this model was 0.98688. For the model trained over the same feature set against 3-month returns, the optimal parameters were an Alpha of 0.27 and an L1 Ratio of 0.01. The minimum MSE for this model was 0.99011.

To select a model to define our composite QARP signal we ran the results of each elastic net in the out-of-sample backtest (described in the following section) in order to select the model with the best test performance regarding Sharpe ratio. We were not surprised to find that the most successful results came from the twelve predictor, three-month window model since the manual predictor selection reduced overfitting and the three-month window was more relevant to the backtest's trading timescale [Fig. 7].

6 Portfolio Construction

In order to test the performance of our QARP signal we sought to perform an out-of-sample backtest over the remaining years of market data (from 2013 to the present). When considering how to construct a portfolio based on this

	Predictor	Signal Magnitude
Value	Book / Market	0.042330506
	Price / Sales	0.015901132
Safety	Receivables / Current Assets	0.000481912
	Operating CF/ Current Liabilities	0.041922074
	Cash Flow / Total Debt	0.013048011
	Cash Balance / Total Liabilities	0.018602827
	Quick Ratio	0.012558456
Profitability	Return on Assets	0.008317245
	Return on Capital Employed	0.009377227
	Return on Equity	0.021725955
	Gross Profit / Total Assets	0.006045664
	Net Profit Margin	0.01141551

Figure 7: Our QARP Signal

signal, we decided to adopt a long-only strategy. This was done partially because the Quantopian backtester compares the result to the S&P 500 ETF (SPY) and partially because Quantopian's rigid Pipeline environment makes market-neutral allocation difficult.

We decided to rebalance the portfolio every three months since company operating information is released on a quarterly basis. Portfolio weightings were assigned through MPT-inspired optimization, which maximized the QARP signal subject to a historical volatility constraint (taken to be maximum risk) available through the "optimization" and "optimization.experimental" classes respectively. We also imposed a constraint of 5% on the maximum size that a position may hold in the portfolio and did not allow for any leverage.

Quantopian accounts automatically for execution costs, in terms of both trading fees and slippage, which supported the real-world applicability of our results. Additionally, the low-frequency nature of the portfolio coupled with the market depth of the quality anomaly indicate that this strategy is relatively robust against execution issues, since trading costs are low and the algorithm is relatively scalable.

7 Results and Discussion

From the elastic net we trained with the top 12 individual fundamentals and with a three-month return window we achieved our best backtesting performance of a 0.76 Sharpe ratio from 2013 to the present, which translated to a p-value of 0.07 statistical significance beyond a null hypothesis of zero. We also beat SPY by 14% during that period and had positive but not statistically significant CAPM alpha [Fig. 8].

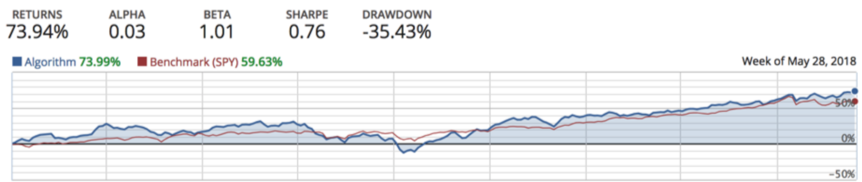

Figure 8: QARP Out-of-Sample Backtest (Bull Market)

We also performed an in-sample backtest that included the Global Financial Crisis and noted the "flight to quality" that is consistently documented during economic downturns [Fig. 9]. This helps confirm that long-QARP strategies pose significantly less drawdown potential than that of the greater market, assuming comparable overall risk exposure.

Figure 9: QARP In-Sample Backtest (Bear Market)

From these results it should be first noted that our portfolio optimization is not a true MPT optimization because the available volatility constraint is simply a weighted average of the historical volatilities of the stocks within the portfolio instead of the overall volatility of the portfolio derived from the asset covariance matrix. [6] This was an unfortunate reality of Quantopian's "optimization.experimental" class that in combination with recent changes to the Pipeline preventing outside optimization packages made achieving a true volatility constraint very difficult if not impossible given the way our pipeline was set-up. [7] Moving forward with this project, the obvious first step would be to correct for this in order to achieve genuine MPT optimization. This would be especially helpful to limit drawdown and relax the maximum position constraint, since properly executed MPT naturally increases the number of securities within a portfolio in order to lower overall volatility. Until that change is made, however, lowering the maximum position constraint is the most direct way to diversify against outsized moves.

Furthermore, a simple, yet likely potent improvement to this project could be an expansion of our dataset to include more companies and more years of data. The inclusion of an industry code would also allow for model segmentation, which might show considerable improvement since fundamental ratios can vary significantly in their value and meaning amongst different industries. More difficult but also potentially beneficial would be a more rigorous treatment of our data before applying linear regression, namely the application of transformations that improve the residual assumptions that underlie linear regression.

While we took the approach of prior literature to find a quality signal whose predictive power persists over time, it is also likely that results could be improved if the fundamental signal constructions are allowed to mutate over time much as the market and its underlying psychology mutates over time. The challenge of developing this dynamic model would be finding enough data to conclude significance in more transient fundamental signals, as well as then finding the right combination with more persistent phenomena.

Finally, a data-mining approach to fundamental selection could uncover previously underutilized metrics if we found access to a much larger universe of ratios. In this case, data snooping could be mitigated through the application of the "reality-check" statistical analysis pioneered in 2000 by Halbert White. [8]

References

[1] William Sharpe. Capital asset prices with and without negative holdings. *Nobel Lecture*, 1990.

[2] Jean-Philippe Bouchaud, Stefano Ciliberti, Augustin Landier, Guillaume Simon, and David Thesmar. The excess returns of 'quality' stocks: A behavioral anomaly. 2016.

[3] Clifford Asness, Andrea Frazzini, and Lasse Pedersen. Quality minus junk. 2013.

[4] John Hoffmann. Linear regression analysis: Assumptions and applications. 2008.

[5] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. 2005.

[6] Harry Markowitz. Portfolio selection. 1952.

[7] Quantopian. Quantopian risk model.

[8] Halbert White. A reality check for data snooping. 1952.