2021

How Good are Equity Analysts? Investigating the Impact of Analyst Recommendations on Portfolio Performance

Maxwell Dawson

Claremont McKenna College

How Good are Equity Analysts? Investigating the Impact of Analyst Recommendations on Portfolio Performance

submitted to
Professor Benjamin Gillen

by
Maxwell Dawson

for
Senior Thesis
Fall 2020
November 30, 2020

Abstract

This paper lies at the convergence of the portfolio optimization literature and the equity research industry. I attempt to quantify the benefit provided to an investor by equity analysts from an asset allocation perspective, and hypothesize that no significant benefit exists because of the incentive misalignments facing analysts as well as the inherent difficulty of valuing stocks. Using a Bayesian regression framework with prior beliefs for alpha generated based on equity analysts' recommendations, I find no significant difference in out-of-sample performance between portfolios which take analyst recommendations into account and portfolios which do not.

Introduction

On March 2, 2017, shares of Snap Inc. (SNAP) began trading publicly on the New York Stock Exchange for $17 a share. At a roughly $33 billion implied equity valuation, the IPO placed the newly public SNAP on the level of well-known giants like Target and Marriott. Morgan Stanley, the leading investment bank on the transaction, waited until March 27 to issue its buy rating and corresponding $28 price target, implying more than 60% upside from the IPO price. Just a day later, on March 28, the bank issued a second note which contained, at the bottom of the second page, the following disclaimer:

> We have corrected a tax calculation error in our model that overstated adjusted EBITDA in 2021-2025. We have updated the text and charts in the following note to reflect our estimate changes. Note that our revenue forecast and fundamental top-line drivers (DAUs, ad load, etc.) remain unchanged.

The overstatement of adjusted EBITDA mentioned here amounted to a decrease in Snap's forecasted free cash flow of over $900 million per year, on average – a 35% decrease. And yet, Morgan Stanley's second note went on to reiterate the bank's $28 price target. By concurrently lowering the cost of capital used in the model from 9.7% to 8%, the bank ended up almost exactly offsetting the overstatement of EBITDA. This change placed Morgan Stanley's cost of capital well below the estimates of other banks, which used costs of capital of anywhere between 11% and 16%.[1]

At best, the research team responsible for constructing this model clearly made two significant errors – by the note's own admission. At worst, "it almost feels like they're backing into the numbers," according to Charles Lee, professor at Stanford Graduate School of Business.[2] Either way, the correction calls into question the accuracy and reliability of analyst price targets

[1] https://finance.yahoo.com/news/morgan-stanley-made-error-analyzing-195438221.html.
[2] Ibid.

in general. While the majority of the extant literature on this subject analyzes analyst rating changes as a near-term cause for returns, it does not address the realized accuracy of the analysts' recommendations, or comment on whether they add value to an investor seeking to generate strong risk-adjusted returns. This paper attempts to answer that question.

I combine a Bayesian regression framework, using analyst recommendation data to create prior beliefs for 12-month forward alpha, with an asset allocation optimization process based on quadratic mean-variance investor utility. By constructing six sets of model specifications, each under different assumptions about the return-generating processes of the included stocks, I directly compare the 12-month returns of portfolios that incorporate analyst recommendations into the Bayesian framework to the returns on portfolios that are agnostic to these recommendations. In light of the increased difficulty of generating alpha in the market due to competition, increased low-cost access to information, as well as the frequent errors and incentive misalignments that plague equity analysts, I hypothesize that portfolios based on analyst recommendations will not outperform portfolios that are not. I first test this hypothesis with a preliminary analysis created by splitting all analyst-rated stocks from a given year into quintiles based on their mean recommendation. This preliminary method shows no statistically significant or economically important difference between the top and bottom quintiles, failing to reject the hypothesis. The full analysis supports the hypothesis as well – across three benchmark models and three analyst-based models, I do not find statistically significant evidence that any of the analyst-based portfolios outperform the benchmark portfolios.

Literature Review

This paper lies at the intersection of the equity research industry and the classic asset allocation problem. As such, this review begins with an overview of the equity research industry and its current challenges. I then present the history of the asset allocation problem, focusing on both the underlying economic theory and the implementation of a utility-maximizing algorithm which performs well both in and out of sample.

For as long as there has been publicly traded equity, there has been equity research. While the nature of the work required to conduct this research has changed over time, the economic niche filled by the equity research industry has stayed largely consistent. Equity analysts, often affiliated with large financial institutions, devote significant resources to valuing publicly traded stocks. By conducting due diligence calls, attending conferences, creating complex models, collecting proprietary data, and more, analysts issue price targets for any number of stocks. These targets are subject to revision at any time, depending on countless factors. The targets, and the accompanying equity research reports, are subsequently used by institutional investors, asset managers, and even some individuals to make investment decisions. The obvious assumption made when an investor integrates one or more equity analysts' conclusions into their decisions is that the analyst has superior analytical tools or information and is therefore able to arrive at a more reliable price target for a given stock. However, as information has become more readily available – annual SEC filings are accessible online, data sources like Bloomberg or Capital IQ are purchased by practically all professional investors, and quarterly earnings calls are transcribed and uploaded to the Internet – investors have become more able to conduct much of this analysis themselves. The question this paper attempts to answer is, in light of these varying complicating factors, to what extent do equity analysts still hold an edge over the rest of the market?

Compounding the increased accessibility of information to unaffiliated investors is the idea that equity analysts may not themselves be objective information aggregators. As an example, Green, Hand, and Zhang (2016) evaluated 120 DCF models produced by sell-side analysts from 2012 and 2013. The authors discovered that analysts made a median of three theory-related or execution errors (e.g. using the wrong risk-free rate) and four questionable economic judgments (e.g. using a market risk premium in excess of 9%) per model. Furthermore, correcting these errors altered price targets by between -2% and 14% per error. Given the median of a combined seven errors per model, the potential for a price target that differs significantly from true value is high. These errors add to, and interact with, the numerous other more subjective assumptions that analysts must make in constructing a model – growing or shrinking margins, revenue growth figures, future capital structure, and more. Green, Hand, and Zhang supplemented their work with interviews with analysts, and came to the following conclusion:

> Based on face-to-face interviews with analysts and those who oversee them, we conclude that analysts‘ DCF modeling behavior is semi-sophisticated in the sense that analysts genuinely make mistakes regarding certain aspects of correctly valuing equity but also respond rationally to the incentives they face, particularly the reality that they are not directly compensated for being textbook DCF correct.

This hints at a further obstacle that analysts face for providing accurate and unbiased price targets: their incentives simply do not encourage them to do so. Many large financial institutions which employ equity analysts also oversee an investment banking division. These divisions are almost always separated by a so-called "Chinese wall," an informational barrier designed to keep the material, non-public information ("MNPI") necessary for investment bankers away from equity analysts. Similarly, the wall is beneficial for insulating equity analysts from outside pressures. To see why, one need only consider a public company deciding which investment bank to engage for a sale. If the equity research division were to raise its price targets for the firm in advance of the announcement, it would likely be able to artificially inflate the price paid by the buyers: and, as a consequence, the fee received by the bank.

Aside from the fact that such informational barriers are rarely perfect (O'Brien, McNichols, and Lin 2005), they are also subverted by many equity analysts' compensation packages (Groysberg, Healy, and Maber 2011). At the bottom of many equity research reports is a disclaimer stating that the authoring analyst's compensation is based in part on the revenue of the entire firm they work for – including the investment banking division. Since investment bankers often charge a fee based on transaction size, the incentive for analysts to try to inflate their price targets and create larger deals for their firm is clear. Between the errors made by analysts in modeling, the incentive misalignments faced by these analysts, and the sheer difficulty of arriving at an accurate price target notwithstanding these two other factors, it is clear that the equity research industry's structure is far from perfect. That being said, its analysts have access to significant resources, information, and opportunities that many other investors do not by virtue of covering only a few stocks. The degree to which the latter mitigates the former in practice is the subject of this paper.

Some authors in the literature have tried to "analyze the analysts," much as this paper does. Jegadeesh et al. (2004) find that sell-side analysts generally recommend stocks that are "positive momentum, high growth, high volume, and relatively expensive" – glamour stocks, as the authors call them. Hong and Kubik (2003) take a different approach, relating analyst forecasts to their career paths. The authors find that analysts with more accurate forecasts are more likely to experience positive career moves, which is to be expected. However, they also find that, independent of the effect of accuracy, analysts who are more optimistic tend to experience more positive career moves (i.e. moving to a higher-status employer). Guttman (2010) endogenizes the timing of analyst forecast issuances, and finds that analysts whose private analysis has a higher degree of precision are more likely to issue their forecasts earlier.

While the literature analyzing the circumstances and consequences of analyst forecasts is somewhat large, this paper distinguishes itself in a few important ways. First, much of the existing literature focuses on the accuracy of analysts' forecasts rather than recommendations – here, I use only recommendation data. Second, the majority of the literature on the subject is devoted to finding and analyzing factors which are likely to determine analyst forecasts. This paper reverses the implied predictive relationship and uses analyst recommendations as an input, rather than an output, to the model to quantify the recommendations' effect on returns.

Of course, quantifying the benefit that equity analysts provide is difficult without a tangible comparison point. This paper uses a combination of Bayesian statistics and the portfolio optimization literature to achieve that goal. Work on the asset allocation problem tends to fall into economic theory, which focuses on finding the return-generating process for assets and the assumptions required to arrive at an optimal portfolio, or implementation, which focuses on the mathematical optimization process.

The most obvious example of the latter is the work of Markowitz (1952), which established the mean-variance optimization framework for quadratic utility. The fundamental problem with this model, which the academic community quickly realized, was its sensitivity to estimation error. While the Markowitz method would produce the optimal portfolio in-sample, it often gave extreme long and short positions in assets that created undesirable levels of volatility out-of-sample. DeMiguel, Garlappi, and Uppal (2009) explored this volatility issue by establishing a "naïve diversification" strategy as the baseline, wherein 1/N of an investor's wealth would be invested in each of N assets. The authors compare a number of asset allocation models with the naïvely diversified portfolio out-of-sample, and show that in many cases the 1/N strategy actually outperforms rigorous mean-variance allocation models. Interestingly, the authors showed that for a 25-asset universe, about 3000 months (250 years) of data is required before a sample-driven optimization algorithm consistently outperforms the naïve diversification. This is obviously an unrealistic amount of data, particularly if one needs to assume it was all generated under the same conditions.

In an effort to resolve this consistently observed volatility in portfolio weights, a different sphere of the literature approaches the problem from a theoretical perspective, introducing factor models to gain a better understanding of the return-generating processes of various securities. The economic theory of portfolio allocation as we know it today began with the foundational work of Sharpe (1963), who established the idea of a factor which drove every stock's return. This gave the area such fundamental concepts as betas and market risk premiums, resulting in the one-factor CAPM often used today. Merton (1969) and Samuelson (1975) then expanded on this, deriving a unique optimal portfolio of risky assets that is independent of an investor's wealth (Merton) or changing risk aversion (Samuelson). Combined, these contributed to the existence of a unique "tangency portfolio" dependent on the risk-free rate – the optimal set of risky assets to hold, regardless of an investor's risk preferences and only assuming a quadratic utility function of the form $U = E[R] - \frac{1}{2}\gamma\sigma_R^2$. Subsequently, however, the academic community discovered that asset returns deviated from the CAPM in predictable ways. The most frequently cited example of this is Fama and French's (1993) three-factor model, wherein asset returns are dependent on market returns as well as relationships to a size portfolio and a value portfolio. The desire to better explain the cross-section of expected returns has contributed to a surge in the number of potential factors surveyed, as well as in the way these factors are examined. The more recent work of MacKinlay and Pástor (1999), for instance, examined the portfolio allocation results in a market that is aware of all but one factor that influences expected returns. Though the paper makes some strong assumptions about the nature of observable factors as well as the idiosyncratic risks of various assets, it is a good example of the way that the factor model was expanded after William Sharpe's initial work.

While the portfolio optimization approach takes volatile initial weights as a given and attempts to mitigate them, the theoretical approach tends to try to quantify the way the market functions. The natural inclination, then, is to merge these two approaches in search of a model that takes in market data, but allows it to be altered in controlled ways to avoid the extreme allocations of Markowitz's model. It is for this reason that the model of Black and Litterman (1992) is perhaps a more promising angle to arrive at a portfolio that consistently performs well out-of-sample. The model takes a quasi-Bayesian approach, starting with the market weights for various assets and then "tilting" them based on the strength and type of an investor's beliefs. The model accounts for absolute beliefs (e.g. "industrials will return 9% this year") as well as relative ones (e.g. "consumer staples will outperform consumer discretionary by 2% this year"), making it highly reflective of the way investors often contemplate their own investment decisions. By construction, the model produces portfolio weights that are reasonably close to the market weights, deviating only in ways that directly contribute to the investor's prior beliefs.

As with most Bayesian models, however, the question then becomes how to source the prior beliefs and the strengths of these beliefs. Pástor and Stambaugh (2002) used an asset-pricing model to generate priors for equity mutual funds, and even with a moderate amount of confidence the results show that including a pricing model meaningfully increases the Sharpe ratio of the resulting portfolios. Bartram and Grinblatt (2018) take a similar approach, using market replicating portfolios of a given firm's financial statements to estimate a "peer-implied fair value" for each security. By finding linear combinations of other publicly traded firms' financial statements that mirror the target's financial statements, Bartram and Grinblatt generated a target price by calculating the actual market price of the linear combination. While these authors simply use the peer-implied fair values to trade rather than including them in a Bayesian asset allocation model, their strategy saw significant success. Gillen (2016) adds the technique of subset optimization to this area, a method which can theoretically be applied to any multi-asset optimization problem. Rather than calculating the weights of all N assets simultaneously, subset optimization randomly generates M portfolios each composed of $\hat{n}$ assets, where $M\hat{n}$ is much larger than N. After optimizing within each subset portfolio, the overall portfolio is then composed of the equally weighted portfolio of all the subsets. Gillen shows that this dominates both the sample-based optimization algorithm of Markowitz and the 1/N rule examined by DeMiguel, Garlappi, and Uppal (2009).

In this paper, I combine the conceptual bases of Pástor and Stambaugh (2002) and Black and Litterman (1992) with both the method of Gillen (2016) and a set of prior beliefs based on equity analyst expectations. The three benchmark portfolios created provide portfolio weights based solely on historical return-generating processes with a prior expectation of alpha being zero, while the experimental portfolios provide updated sets of weights where the prior estimates of alpha are based on equity analyst predictions. Consequently, a comparison of these portfolios and their performance will yield a clear picture of the exact ways in which the inclusion of equity research consensus estimates impacts the portfolio performance.

Data

I use two databases in the construction of the sample for this paper. Historical stock return data are from CRSP, and equity analyst recommendation data – including the number, mean and standard deviation of recommendations – are from Thomson Reuters' I/B/E/S. A recommendation in this context is a rating by a single analyst of a single stock, issued as an integer between 1 and 5 with 1 indicating a "strong buy" stock and 5 indicating a "strong sell" stock. The I/B/E/S dataset, containing data back through December 1993, is the lower limit of the sample, with the upper limit being December 2018 I/B/E/S data for a total of 26 years. However, only each December is used in order to maximize theoretical predictive power for the following year's returns. For each year N of analyst recommendation data, I take the 11 years ranging from $N-9$ to $N+1$ of stock returns from CRSP. This is why the 2019 I/B/E/S data are not used – 2020 CRSP data are not yet available as of the writing of this paper. The CRSP data consist of monthly stock returns for each year, for a total of 132 observations for each stock which trades for the full 11 years. For stocks which are delisted during the period, the return in the month of delisting represents the final return observed for that stock.

For a given year, the first ten years of CRSP data are used to calibrate the Bayesian model and produce estimates of the moments of the data. The final year of CRSP data functions as the out-of-sample test period. Since many price targets produced by analysts are designated as 12-month price targets, additional out-of-sample years would not relate to the stated goal of many analysts' recommendations. The sample consists of the intersection of the CRSP and I/B/E/S databases (i.e. all stocks which had at least one analyst issuing recommendations) for each year. The I/B/E/S and CRSP datasets are merged on CUSIP, a unique identifier shared by both sources. However, some stocks within the I/B/E/S data for various years do not have a

CUSIP attached to them. These observations are dropped from the sample for that year, but do not exceed 2% of the sample in any year so it is unlikely these affect the data significantly. Table 1 below shows the number of companies included in the sample for each year.

Table 1: Number of stocks included in out-of-sample period for each year

1994	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
774	924	941	991	1048	1039	1032	986	971	1046	1203	1314	1407

2007	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
1470	1475	1531	1651	1694	1739	1775	1773	1811	1837	1882	1916	1894

Additionally, some stock returns are missing from the CRSP database for a given month; these returns are replaced with the S&P500 return for the same month, when they fall in the out-of-sample period of a given year's data. After all these adjustments and merges, the final dataset used for one year of the analysis herein consists of mean ($meanrec$), standard deviation ($stdev$), and number of recommendations ($numrec$) for all stocks in a given year, with the ten years of returns before and one year after ($returnYYYYMMDD$). 26 years sampled in this fashion give the complete dataset utilized in this paper.

Table 2: Summary statistics for I/B/E/S data from selected years in sample

	meanrec	***stdev***	***numrec***
1994 (n = 774)			
Mean	2.28	0.63	4.95
Median	2.27	0.71	3
Standard deviation	0.74	0.52	4.19
2006 (n = 1,407)			
Mean	2.45	0.70	8.43
Median	2.47	0.79	6
Standard deviation	0.65	0.45	7.42
2019 (n = 1,894)			
Mean	2.32	0.67	9.90
Median	2.29	0.74	7
Standard deviation	0.56	0.37	8.31

A few trends are immediately notable from the basic summary statistics of the data. First, as is perhaps expected, the number of stocks in the sample generally rises as time goes on. This is likely due both to an increase in trading activity and the increase in prevalence of analysts over the period. The latter is also seen clearly in the summary statistics for $numrec$ – such that the mean number of recommendations for a given stock in the sample almost exactly doubled from 1994 to 2019. As would be expected, the median value also rose, though it is consistently lower than the mean. This occurs because, while $numrec$ has a floor at 1 in the sample, a few blue-chip stocks have upwards of 40 analysts covering them in 2019. The summary statistics of $meanrec$ are perhaps the most notable, with means consistently very similar to medians. Two trends, however, are clear: first, both the mean and median of $meanrec$ are consistently less than 3. This is reflective of a tendency of analysts to preferentially rate stocks as buys over sells. It does appear irrational that analysts overall expect the average stock to outperform the market – but this is perhaps to be expected in light of the various incentive misalignments facing equity analysts, as previously discussed. Furthermore, the standard deviation of $meanrec$ is steadily dropping over the period, indicating that analysts are increasingly hesitant to issue strong ratings of 1 ("strong buy") or, even more likely, 5 ("strong sell"). The summary statistics of $stdev$ have a less clear mapping to real-world effects, but it is interesting to note that the standard deviation of $stdev$ is declining as well. This is suggestive of the idea that analysts have tended to disagree with each other in increasingly similar ways on different stocks. Though there is not an immediately evident real-world explanation for this, it is an interesting trend nonetheless.

Preliminary Analysis

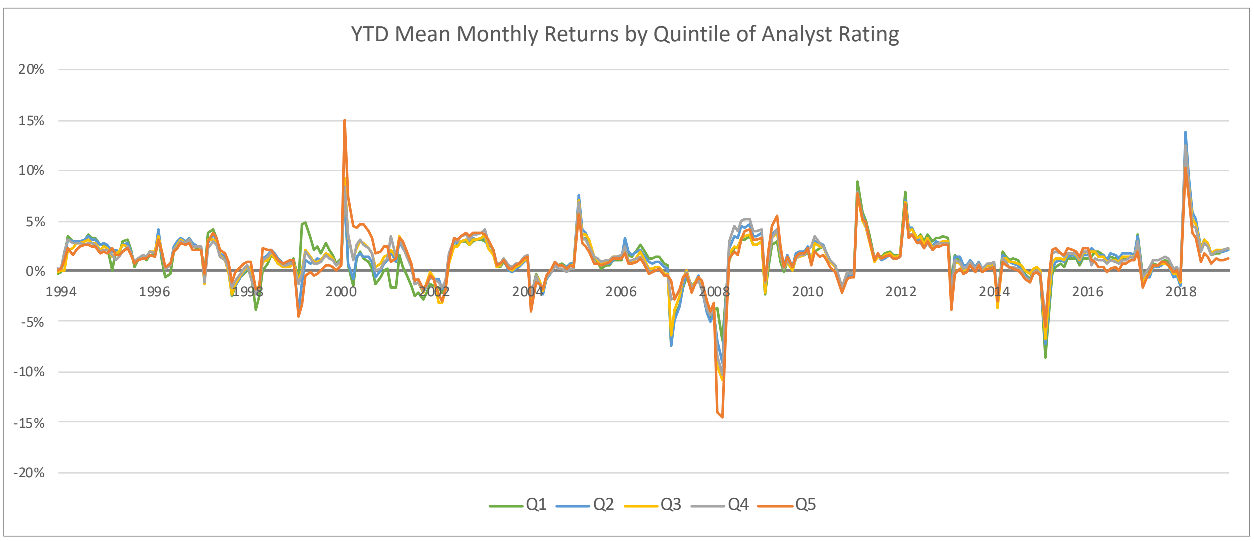
As a preliminary analytical tool, and to create a more intuitive understanding of the data, the stocks from a given year were separated into quintiles based on their mean ratings and their out-of-sample returns were compared. To implement this, a standardized metric $T_{i,t}$ resembling a t-statistic was constructed for each stock in each year using the following formula:

$$T_{i,t} = \frac{meanrec_{i,t} - \overline{meanrec_t}}{\frac{stdev_{i,t}}{\sqrt{numrec_{i,t}}}}$$

In this construction, $meanrec_{i,t}$ represents the mean analyst recommendation for stock i in year t; $\overline{meanrec_t}$ represents the average mean recommendation over all stocks in year t; $stdev_{i,t}$ represents the standard deviation of recommendations for stock i in year t; and $numrec_{i,t}$ represents the number of recommendations for stock i in year t. In effect, this is a standardized measure of a stock's rating relative to the rating for all stocks in the sample for that period. For some stocks, $stdev_{i,t} = 0$; this occurs when $numrec_{i,t} = 1$ or when every covering analyst for the stock rated it the same. In these cases, the data are winsorized so that the 5th percentile of the nonzero observations of $stdev_t$ is substituted for $stdev_{i,t}$. This allows a value of $T_{i,t}$ to be calculated for these observations, while staying approximately faithful to the true observed standard deviation of 0.

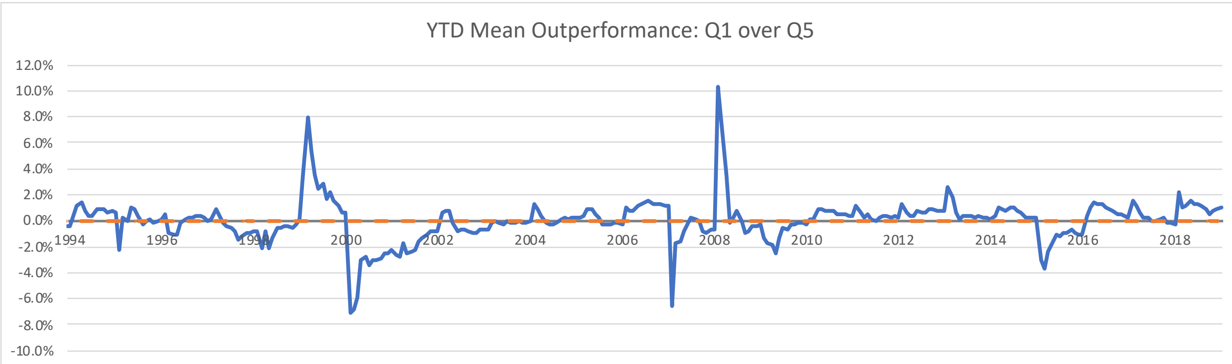
The stocks for each year are then divided into quintiles based on their values for $T_{i,t}$ – the lowest values of $T_{i,t}$, and thus the first quintile, corresponding to the stocks most favorably rated by analysts. Monthly quintile returns for a given year are then calculated as the return on an equally-weighted portfolio of all assets in the quintile for that month. Repeating this process for all years in the sample produces a continuous set of monthly returns from 1994 to 2019 for each quintile, with each year's returns representing the returns on the stocks rated in that quintile at the end of the prior year. For a given month – without loss of generality, I use April 2019 here as an illustrative case – I then calculate the year-to-date average of returns. For April 2019, this is the geometric mean of the returns from January 2019 to April 2019. Since the quintiles for 2019 returns are generated based on December 2018 analyst recommendations, the year-to-date geometric average monthly return functions as a direct measure of returns over the period the analyst recommendations are intended to predict. Chart 1 below shows the YTD average return of each quintile over the entire sample period.

Chart 1: Year-to-date mean monthly returns for each quintile of analyst rating



In the chart above, there appears to be little consistent difference between Q1, the most favorably rated stocks, and Q5, the least favorably rated stocks. However, to statistically confirm this, I adopt the null hypothesis that there is no difference between YTD average returns of Q1 and Q5 with the one-sided alternative that Q1 outperforms Q5. Chart 2 below plots the outperformance of Q1 over Q5, with the dotted orange line representing the average outperformance over the sample period.

Chart 2: Year-to-date mean monthly outperformance of most positively-rated quintile over most negatively-rated quintile



The initial appearance of no significant difference between Q1 and Q5 is confirmed by a difference in means test – on average, Q1 outperforms Q5 by just 1.3 bps (9.2 bps) per month, giving a t-score of 0.14 which is insufficient to reject the null at any conventional level. One complicating factor is the few extreme values observed, for instance in the first month of 2008 when Q1 outperforms Q5 by more than 1,000 bps. Most of the extreme values visible in the chart are due to the nature of YTD averaging. In the earliest months of a given year, there are fewer months of returns contributing to the average – as such, the YTD average will have an unusually high standard error, indeed equaling the volatility of a single month's returns in the first month. By December, however, volatility in any one month is unlikely to affect the average excessively. One way to mitigate this is to consider the same test above, but with YTD averages from the first quarter of the year removed. Over this sample, Q1 actually underperforms Q5 by an average of 4.4 bps (6.6 bps). While the removal of means from the first quarter is effective in lowering volatility by 28%, the results still yield a t-score of -0.67 – again not significant at any conventional level. Additional tests on removing the second and third quarters are shown in Table 3 below.

Table 3: Effects of removing first *n* quarters of returns on statistical significance of top-quintile outperformance over bottom quintile			
Test	**Mean**	**Standard error**	**T-score**
All Quarters Included	1.3 bps	9.2 bps	0.14
First 1 Quarter Removed	-4.4 bps	6.6 bps	-0.67
First 2 Quarters Removed	-6.4 bps	6.9 bps	-0.92
First 3 Quarters Removed	-8.6 bps	8.6 bps	-1.00

It is clear from the table above that removing additional quarters beyond the first is not effective; indeed, it actually increases the standard error due to the substantial decrease in sample size that removing these additional quarters causes. These results suggest that the most favorably rated stocks do not significantly outperform even the least favorably rated stocks in the year after they are rated; put simply, they imply that analyst ratings are not a reliable predictor of returns in the 12 months after the ratings are given.

Method

Beginning from the merged CRSP-IBES dataset discussed above, the method of analysis I conduct here can be summarized in four steps. First, the prior estimates of alpha for each stock are generated. Second, Fama and French's factor data are imported and each stock's returns are regressed on these variables, which are then used in the standard Bayesian process to obtain the posterior expectations and variances of returns for each security. Third, the process of subset optimization discussed by Gillen (2016) is used to generate the optimal portfolio weights for each period. Finally, these weights are applied to the out-of-sample return periods to build the out-of-sample performance. Each of these steps is discussed in detail below.

The prior estimates of $E(\alpha)$ are generated on a linear scale based on each stock's mean recommendation in comparison to the average mean recommendation across all assets for the period. The stock with the value of $meanrec$ which is furthest from $\overline{meanrec}$ over the period will be assigned the greatest magnitude of α. Specifically, the estimate for stock i is generated with the following formula:

$$E_{prior}(\alpha_i) = -\frac{1}{12} * \alpha_{max} * \frac{meanrec_{i,t-1} - \overline{meanrec_{t-1}}}{\max(|meanrec_{t-1} - \overline{meanrec_{t-1}}|)}$$

$meanrec_{i,t}$ is the same variable discussed previously, representing the mean analyst recommendation for stock i at time t; α_{max} represents the maximum value I allow the prior estimate of annual alpha to take. The selection of α_{max} varies in the model based on the confidence I place in the analyst estimates; values of 0%, 1%, 5%, and 10% are used here. The scaling factor of $-1/12$ converts this figure to a monthly parameter to match with the monthly stock returns used in the data, and the negative sign accounts for the fact that a $meanrec$ of 1 corresponds to a strong buy, and therefore should have positive alpha. It should be noted here that, since $\overline{meanrec}$ is typically less than 3, the stocks with the largest magnitude of alpha will typically be those rated as strong sells, with consequently large negative alpha. The prior estimate for the standard deviation of alpha is generated in a similar way:

$$\sigma_{prior}\left(\alpha_{i,t}\right) = \frac{1}{12} * \alpha_{max} * \frac{stdev_{i,t-1}}{\max(|meanrec_{t-1} - \overline{meanrec_{t-1}}|)} + \sigma_\alpha$$

As in the typical Bayesian formulation, the first term of this expression corresponds to the standard deviation of the recommendations for stock i, scaled by the same factor used above. Though this is a standard deviation and not a variance, the scale of $1/12$ is still appropriate since its only purpose is to scale α_{max} to a monthly level. The second term, σ_α, is a parameter set in the model, and represents the expected value of the standard deviation of α independent of the variance in expectation calculated in the first term. Values for σ_α of 0.0001%, 1%, and 100% are used in this paper, and each corresponds to a distinct model which will be discussed in more detail later.

After generating the priors, the next step is to bring in the Fama and French data. This paper uses six of the Fama-French factors: market performance, size, value, momentum, short-term reversal, and long-term reversal. I first convert the sample returns to excess returns by subtracting the risk-free rate for the corresponding period, then regress these excess returns on the benchmarks to pull the residuals. The mean squared residual across all assets and all time periods within the in-sample period is then used to scale the prior variance. I then use the standard Bayesian shrinkage process to arrive at the posterior expected returns and variances for each security.

The following step is to use these posterior expected returns and variances as inputs into the subset optimization process of Gillen (2016). I employ two constraints – the "full investment" constraint that portfolio weights must sum to one, and the "no short-selling" constraint that no portfolio weights can be less than zero. A corollary of these rules is that the portfolio may not be more than 100% invested in any asset. I use the standard quadratic utility formulation of investor preferences, where $U = E(R) - 0.5\gamma\sigma_R^2$. γ, representing the investor's risk aversion coefficient, is set to 4 for this paper. I take 1000 subset portfolios, each with 50 stocks. Each subset portfolio is generated with a random selection of 50 stocks (with replacement), and the subset is then optimized using quadratic utility. After the formulation of all subset portfolios is complete, the overall optimal portfolio is composed of the equally-weighted portfolio of these subsets. Once the optimal portfolio weights are calculated, it is fairly simple to test the portfolio on the out-of-sample returns. However, note that each out-of-sample period is only one year. This entire process, then, is conducted 26 times – one for each sample period – which results in an optimal portfolio that is rebalanced each year. For the purposes of this paper, the transaction costs of rebalancing a portfolio which may contain nearly 2000 assets are not accounted for.

Results

In the preliminary analysis, it was shown that no significant difference existed between the return on the top analyst-rated quintile of stocks and the bottom analyst-rated quintile. Consequently, both the hypothesis and that analysis suggest that the portfolio allocation approach is unlikely to yield any significant difference between portfolios which take analyst expectations into account and portfolios which do not. To test this, I analyze six portfolio construction strategies which vary in their values of α_{max} and σ_α:

1) **CAPM: $\boldsymbol{\alpha_{max} = 0\%, \sigma_\alpha = 0.0001\%}$**

The first model I construct is the standard six-factor CAPM, with a maximum prior expectation for alpha of 0% and an arbitrarily small prior standard deviation of alpha. This produces an allocation strategy which assumes the six factors considered – market, size, value, momentum, short-term reversal, and long-term reversal – explain the systematic return-generating process completely and alpha does not exist. Note that setting $\sigma_\alpha = 0$ would lead to perfect singularity in the matrices used to execute the Bayesian shrinkage process, and I therefore use the arbitrarily small 0.0001% instead to ensure the extremely high precision of this estimate causes it to dominate any alpha seen in the data when the posterior estimate is calculated. The standard CAPM serves as one of the three benchmark models used.

2) **Bayesian CAPM: $\boldsymbol{\alpha_{max} = 0\%, \sigma_\alpha = 1\%}$**

The second model considered is a Bayesian model of the six-factor CAPM, with a maximum prior expectation for alpha of 0% and a prior standard deviation of 1%. While this is still an uninformed prior expectation like model (1), it allows for some alpha to be introduced from the data. Stocks that appear to generate alpha over the training period will be more strongly weighted in this model than in model (1). The Bayesian CAPM serves as the second of the three benchmarks used.

3) Agnostic Bayesian CAPM: $\boldsymbol{\alpha_{max} = 0\%, \sigma_\alpha = 100\%}$

The third model considered is another Bayesian model, with a maximum prior expectation for alpha of 0% but an arbitrarily large standard deviation. The effect this produces is that, because the precision of the prior estimate of alpha is so low, any small amount of alpha seen in the data will dominate the prior expectation in the posterior estimate. Consequently, while model (1) assumes the investor is confident alpha cannot be generated, and model (2) assumes the investor believes alpha cannot be generated but could learn otherwise from the data, model (3) effectively assumes the investor knows nothing at all about alpha before incorporating the data. Because of this juxtaposition of three benchmark models, we are able to clearly see whether any kind of investor would be able to benefit from incorporating analyst recommendations: independently of their beliefs about alpha.

4) Analyst 1%: $\boldsymbol{\alpha_{max} = 1\%, \sigma_\alpha = 1\%}$

The fourth model considered is the first one which incorporates analyst expectations, with a maximum prior expectation of alpha of 1% and a standard deviation of 1%. Using the process discussed earlier, this model standardizes the prior belief for alpha on the interval $[-1\%, 1\%]$, with the highest-rated stocks receiving positive alpha and the lowest-rated stocks receiving negative alpha. While this model (like the next two which also incorporate analyst expectations) allows for shrinkage based on the data, it creates shrinkage towards a different value.

5) Analyst 5%: $\boldsymbol{\alpha_{max} = 5\%, \sigma_\alpha = 1\%}$

The fifth model considered is exactly the same as the fourth, but with a broader interval for the prior estimate of alpha to fall in – namely the interval $[-5\%, 5\%]$. By using the same standard deviation with a larger range, this model produces an effect of greater confidence in the analysts' recommendations as a predictor of returns. Because the precision of the estimate remains the same, values of alpha in this model will be more extreme than those in model (4).

6) **Analyst 10%: $\boldsymbol{\alpha_{max} = 10\%, \sigma_\alpha = 1\%}$**

The sixth and final model is the same as models (4) and (5), with an even larger interval of possible alpha priors: $[-10\%, 10\%]$. Since this model again has the same standard deviation as models (4) and (5), it only produces a greater spread of the posterior alpha results without decreasing the precision of the prior. It is also worth noting that 10% was the largest value of α_{max} used because it is unlikely any investor would consistently expect more than 10%, or less than -10%, annual alpha from a single stock. The summary statistics of the excess returns of the six portfolios generated by these models are in Table 4 below.

Table 4: Summary statistics of portfolio performance under each of six sets of model specifications

Model	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	CAPM	**Bayesian CAPM**	**Agnostic Bayesian CAPM**	**Analyst 1%**	**Analyst 5%**	**Analyst 10%**
Parameters						
$\boldsymbol{\alpha_{max}}$	0%	0%	0%	1%	5%	10%
$\boldsymbol{\sigma_\alpha}$	0.0001%	1%	100%	1%	1%	1%
Summary Statistics						
Mean	0.92%	0.90%	0.85%	0.90%	0.90%	0.92%
Standard deviation	4.80%	4.75%	5.40%	4.74%	4.71%	4.70%
Annualized mean	10.99%	10.74%	10.20%	10.78%	10.83%	11.05%
Annualized SD	16.63%	16.45%	18.70%	16.40%	16.30%	16.29%
Sharpe ratio	0.661	0.653	0.545	0.657	0.664	0.678

With the exception of model (3), the agnostic Bayesian CAPM, all the models appear to have almost identical Sharpe ratios. This is driven by largely similar means and standard deviations, except for model (3) which has a notably higher standard deviation of returns. A formal set of tests on the monthly outperformance of each model over the three benchmarks confirms this initial impression, the results of which are shown in Table 5 below.

Table 5: Outperformance statistics of three analyst models over three benchmark models

Model	**(4)** **Analyst 1%**	**(5)** **Analyst 5%**	**(6)** **Analyst 10%**
Model (1): CAPM			
Mean outperformance	-1.8 bps	-1.4 bps	0.4 bps
Standard error	2.7 bps	3.1 bps	3.8 bps
T-Score	-0.66	-0.44	0.11
Model (2): Bayesian CAPM			
Mean outperformance	0.3 bps	0.7 bps	2.6 bps
Standard error	0.7 bps	1.5 bps	2.5 bps
T-Score	0.47	0.49	1.03
Model (3): Agnostic Bayesian CAPM			
Mean outperformance	4.8 bps	5.2 bps	7.1 bps
Standard error	8.6 bps	8.5 bps	8.4 bps
T-Score	0.56	0.62	0.84

Clearly, none of these findings are sufficient to reject any null hypothesis that models (4), (5), or (6) do not outperform the benchmark models (1), (2), or (3) at any conventional level of significance. As was expected from the preliminary analysis, there is no evidence here to suggest that incorporating analyst recommendations into the portfolio allocation process has any positive effect on expected return. We can also examine the return-generating processes of each of the six portfolios, listed below in Table 6.

Table 6: Return-generating processes for each of the six model specifications tested

Model	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	CAPM	**Bayesian CAPM**	**Agnostic Bayesian CAPM**	**Analyst 1%**	**Analyst 5%**	**Analyst 10%**
Alpha	0.21%	0.20%	0.15%	0.21%	0.22%	0.25%
	(0.06%)	(0.06%)	(0.08%)	(0.06%)	(0.06%)	(0.06%)
$\boldsymbol{\beta_1}$**: Market**	0.96	0.96	1.04	0.96	0.94	0.92
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
$\boldsymbol{\beta_2}$**: SMB**	0.39	0.38	0.45	0.39	0.41	0.44
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\boldsymbol{\beta_3}$**: HML**	0.30	0.21	-0.12	0.21	0.20	0.19
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
$\boldsymbol{\beta_4}$**: MOM**	-0.10	-0.07	-0.08	-0.07	-0.07	-0.07
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
$\boldsymbol{\beta_5}$**: ST Rev.**	0.04	0.04	0.03	0.04	0.05	0.05
	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
$\boldsymbol{\beta_6}$**: LT Rev.**	0.01	0.01	0.05	0.01	0.02	0.02
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)

Unsurprisingly, no large difference exists between the alpha estimates of the six sets of specifications. The coefficients on the various Fama-French factors vary slightly from one set of specifications to the next, but there is no significant change between the benchmark portfolios and the analyst-based portfolios.

To confirm the completeness of the six sets of specifications used, a series of robustness checks are employed. The first consists of a seventh set of specifications, where $\alpha_{max} = 10\%$ and $\sigma_{\alpha} = 10\%$ – effectively a combination of the agnosticism from model (3) and the high degree of analyst confidence from model (6). This model produces results almost identical to the results of model (3), with a mean excess monthly return exactly equal to that of model (3) and a standard error just 3 bps lower. Consequently, like model (3), this additional model has a lower Sharpe ratio than both the other two benchmarks and the three analyst-based models: though neither causes statistically significant outperformance. This suggests, albeit not decisively due to the lack of significance, that the agnostic approach (using high σ_α) produces unfavorable results independent of the value of α_{max}. This is perhaps to be expected since such an approach, with a low precision for the prior estimate, negates the benefit of using Bayesian regression.

Two additional robustness checks involve using a value of γ, the risk aversion coefficient, of 1 instead of 4; as well as using 100-stock subset portfolios instead of 50-stock portfolios for the subset optimization. The first, which represents the investor being more willing to take on risk, succeeds in increasing volatility – however, it too fails to increase expected returns and therefore yields an even lower Sharpe ratio than any of the benchmarks. The second similarly yields no significant change in the results, indicating that the initially selected value of $\hat{n} = 50$ creates sufficient diversification within the subsets. While these three checks are not a complete assessment of the possible variations to the model specifications, they each alter a primary input to the portfolios – σ_α, γ, and $\hat{n}$ – and the lack of significant differences suggests the six sets of specifications used are sufficient to address the question.

One final check I employ is to shorten the sample period to the 17 years from 2003-2019. After the passage of the Sarbanes-Oxley Act in 2002, a number of protections were implemented in an attempt to ensure the separation of the investment banking and equity research practices within a financial institution. Therefore, it is possible that the performance of the analyst-based portfolios would improve after 2002 in light of their new incentive structure. Although the analyst-based portfolios do improve, with an average increase in Sharpe ratio of 0.042, the benchmark portfolios see an even greater average increase in Sharpe ratio of 0.057. The result is that, despite the increase in Sharpe ratio for the analyst-based portfolios, they still fail to outperform the benchmarks. Despite the other effects Sarbanes-Oxley had on the equity research industry, I do not find evidence it significantly altered the industry's collective outperformance.

Conclusion

The hypothesis I investigate is that portfolios based on analyst recommendations do not significantly outperform portfolios not based on those recommendations. To evaluate this hypothesis, I use monthly stock return data from CRSP and analyst recommendation data from I/B/E/S to create analysis for the 26 years from 1994 to 2019. With this data, I employ two different methods to test the hypothesis. The first, and more simplistic, method consists of separating all stocks rated in a given year into quintiles based on their mean recommendation and then evaluating the performance of the equally-weighted portfolios within each quintile. The second analytical method employs a Bayesian regression model in which analyst recommendation data is used to generate the prior estimate for alpha, and mean-variance utility-maximizing portfolios are then constructed under different sets of specifications for the confidence in analyst recommendations. In both the preliminary analysis and the primary analysis, the results fail to reject the null hypothesis, since neither analysis provided evidence of statistically significantly different performance between either the quintiles or the portfolios.

The results are consistent with the hypothesis that analysts do not contribute significant value to a portfolio allocation model. It is worth noting one primary assumption made in this paper, on which further research would almost certainly be productive, is that analysts are a homogeneous group. That is, the analysis herein does not distinguish between those at high-status investment banks and those at less credible institutions. Additionally, this paper does not consider stocks which simply do not receive an analyst rating, since these are not contained in the I/B/E/S database. It could be valuable to investigate whether some analysts tend to be more correct than others, as well as how a portfolio composed of "unrated" stocks performs relative to the portfolios assessed here. As this paper stands, however, it suggests more research is needed as to whether equity analysts are truly able to generate an edge. Equity research is a multimillion-dollar enterprise for many investment banks – financed by all manner of institutional investors – but my results suggest that investors may not be any better off because of it.

References

Bartram, S. M., & Grinblatt, M. (2018). Agnostic fundamental analysis works. *Journal of Financial Economics*, 128(1), 125-147.

Black, Fischer, and Robert Litterman. "Global Portfolio Optimization." *Financial Analysts Journal*, vol. 48, no. 5, 1992, pp. 28–43, doi:10.2469/faj.v48.n5.28.

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

Fama, Eugene F., and Kenneth R. French. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, vol. 33, no. 1, 1993, pp. 3–56, doi:10.1016/0304-405x(93)90023-5.

Gillen, Benjamin J. "Subset Portfolios for Asset Allocation" *Caltech Social Science Working Paper 1421*, 2016, https://authors.library.caltech.edu/79336/1/sswp1421.pdf.

Green, Jeremiah, et al. "Errors and Questionable Judgments in Analysts' DCF Models." *Review of Accounting Studies*, vol. 21, no. 2, 2016, pp. 596–632, doi:10.1007/s11142-016-9352-4.

Groysberg, Boris, et al. "What Drives Sell-Side Analyst Compensation at High-Status Investment Banks?" *Journal of Accounting Research*, vol. 49, no. 4, 2011, pp. 969–1000, doi:10.1111/j.1475-679x.2011.00417.x.

Guttman, I. (2010). The Timing of Analysts' Earnings Forecasts. *The Accounting Review*, 85(2), 513-545. Retrieved November 22, 2020, from http://www.jstor.org/stable/20744140

Hong, H., & Kubik, J. (2003). Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts. *The Journal of Finance*, 58(1), 313-351. Retrieved November 22, 2020, from http://www.jstor.org/stable/3094489

Jegadeesh, N., Kim, J., Krische, S., & Charles M. C. Lee. (2004). Analyzing the Analysts: When Do Recommendations Add Value? *The Journal of Finance*, 59(3), 1083-1124. Retrieved November 22, 2020, from http://www.jstor.org/stable/3694731

MacKinlay, A. Craig, and Lubos Pástor. "Asset Pricing Models: Implications for Expected Returns and Portfolio Selection." 1999, doi:10.3386/w7162.

Markowitz, Harry. "Portfolio Selection." *The Journal of Finance*, vol. 7, no. 1, 1952, p. 77, doi:10.2307/2975974.

Merton, Robert C. "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case." *The Review of Economics and Statistics*, vol. 51, no. 3, 1969, p. 247, doi:10.2307/1926560.

O'Brien, Patricia C., et al. “Analyst Impartiality and Investment Banking Relationships.” *Journal of Accounting Research*, vol. 43, no. 4, 2005, pp. 623–650, doi:10.1111/j.1475-679x.2005.00184.x.

Pástor, Ľ., & Stambaugh, R. F. (2002). Investing in equity mutual funds. *Journal of Financial Economics*, 63(3), 351-380.

Samuelson, Paul A. “Lifetime Portfolio Selection By Dynamic Stochastic Programming.” *Stochastic Optimization Models in Finance*, 1975, pp. 517–524, doi:10.1016/b978-0-12-780850-5.50044-7.

Sharpe, William F. “A Simplified Model for Portfolio Analysis.” *Management Science*, vol. 9, no. 2, 1963, pp. 277–293, doi:10.1287/mnsc.9.2.277.

Turner, Matt, and Rachael Levy. “Morgan Stanley Made an Error Analyzing Snapchat, and It Shines a Light on Some Big Flaws in Wall Street Research.” *Yahoo! Finance*, Yahoo!, 4 Apr. 2017, finance.yahoo.com/news/morgan-stanley-made-error-analyzing-195438221.html.