# GENERATING ABNORMAL RETURNS USING CROWDSOURCED EARNINGS FORECASTS FROM ESTIMIZE

Leigh Drogen
Vinesh Jha
24 September 2013

**Abstract**
We examine consensus EPS and Revenue forecasts derived from the crowdsourced community Estimize, and find that they are more accurate than traditional sell side equity analysts' consensus forecasts. We also show that the post earnings announcement drift effect is more pronounced for earnings surprises that are benchmarked against the Estimize community's expectations as compared with Wall Street's expectations. Finally, we demonstrate that a strategy which seeks to exploit the differences between the two communities' expectations prior to earnings dates earns abnormal residual returns, particularly among large cap stocks.

## INTRODUCTION

Estimize is an online community, established in 2011, in which contributors can supply structured financial forecasts. Contributors can be buy side investment professionals, independent researchers, individual traders, or students; the community is split broadly among these groups, with no one group dominating the sample. Because of the diversity of the backgrounds and methodologies underpinning the contributed forecasts, the estimates in the Estimize system represent a distinct alternative to estimates from traditional sources such as sell side equity research desks. In this study we examine some characteristics of Estimize estimates, and examine whether profitable trading strategies can be derived from them.

At the time of writing, the forecasts are quarterly EPS and Revenue estimates on U.S. equities, with plans to expand to other markets and other types of forecasts. There is a long history of academic research on earnings forecasts, the vast majority of which uses professional sell side estimates from data sources such as Thomson Reuters' I/B/E/S. While the emergence of such databases in the 1970's, and their commercialization among data-driven investors in the 1990's, represented a huge step forward in investors' ability to predict corporate earnings, the data suffers from several potential inherent biases.

Professional forecasters' livelihood depends on providing accurate information to buy side clients. However, in surveys of the buy side, forecasting ability is rarely ranked among the most important characteristics of sell side analysts. Analysts are more valued for providing corporate access, detailed research reports, and general industry information [Boni and Womack, 2002]. Services such as Thomson Reuters' StarMine have brought more accountability to the sell side equity research community, but EPS accuracy is at best only one input into an analyst's compensation and prominence.

Furthermore, there exists evidence of behavioral biases in sell side earnings estimates, including herding behavior [Trueman, 1994; Hong et al, 2000]. For example, analysts can be classified as leaders or followers. Any individual forecaster is unlikely to want to make particularly bold estimates away from the consensus, for fear of being exposed as incorrect. Regulatory constraints on sell side analysts, particularly since the financial reforms in the early 2000's, also constrain an analyst's ability to produce timely updates to their research, as several more compliance steps are now involved. As a result, changes in sell side forecasts tend to be fairly gradual.

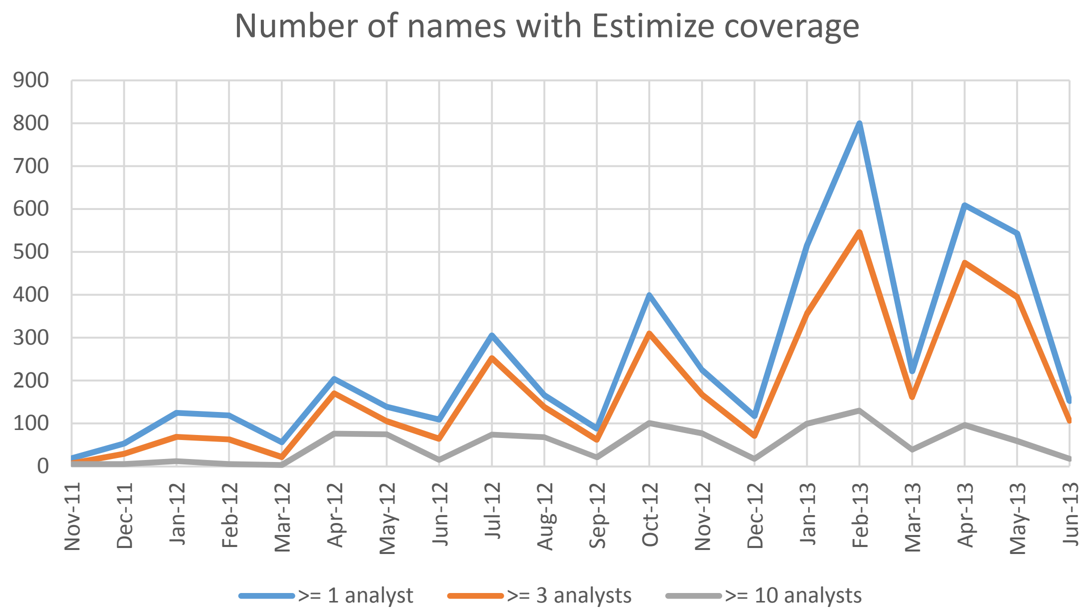
Lastly, there is evidence that institutional biases may exist. Many analysts are employed by large financial institutions which also include investment banking arms. To the extent that these banks want to retain the advisory business of their corporate clients, there may be inherent pressure on the firm to provide unrealistically optimistic forecasts on certain issuers, or at least to avoid pessimistic forecasts [Boni and Womack, 2002; Michaely and Womack, 1999]. In theory there are regulatory practices in place to deter such biases, including required disclosures and information barriers, but these concerns remain.

The Estimize system provides alternative forecasts which may be less affected by such biases. Forecasters on Estimize have the flexibility to provide honest and timely estimates, and are not compensated by the platform for being optimistic or accurate or providing any particular quantity of data. However there do exist incentives to be accurate, including self-promotion, which could be particularly important for independent researchers or students who would like to have a published track record of accurate estimates to point to.

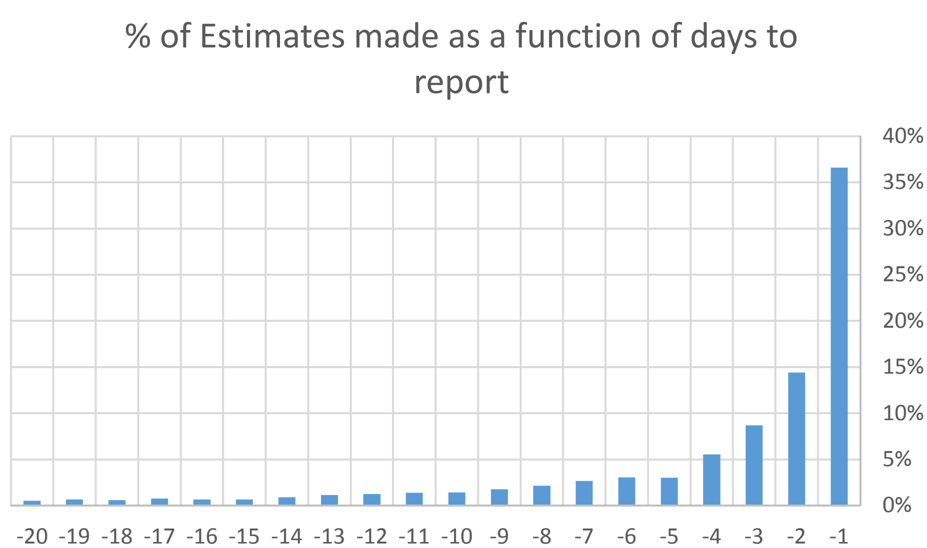
As a crowdsourced platform, Estimize takes advantage of the "wisdom of the crowds" and therefore provides a diversifying source of information. There is a growing body of literature that finds that crowdsourced forecasts can be particularly accurate [Avery et al, 2011; Wolfers and Zizewitz, 2006]. We contribute to such studies by looking at the accuracy of Estimize consensus earnings estimates, and by measuring whether the market response to earnings announcements indicates that the Estimize community is an accurate measure of market expectations. Finally, we examine a trading signal called the Estimize Delta which is based on the deviation between the Estimize consensus and the professional Wall Street consensus, and find that it can generate abnormal returns, particularly among large cap U.S. stocks, indicating that the information in the Estimize consensus has not been arbitraged away.

## DATA

Estimize was founded in 2011; we begin our analysis in November 2011 and end it in July 2013. The Estimize community grew rapidly through 2011 and 2012, although in the early months of our sample the data was fairly sparse. An important caveat in these studies is that with a relatively short historical sample, and with a community whose characteristics and contributor base may evolve over time, we need to be careful about extrapolating our findings into the future.




There is strong seasonality in the data. Most contributors make estimates in the two weeks leading up to an earnings report, and there are relatively few estimates made for quarters beyond the current (FQ1) quarter, though the platform does allow for estimates out to FQ4. As a result, there will be relatively few trading signals that we will be able to calculate in the lull between reporting seasons.




Crowdsourced estimates by their nature can be fairly noisy. Estimize attempts to mitigate the noise in two ways. First, when a contributor enters an estimate into the system, the "starting point" estimate is the Wall Street consensus, and that value is displayed next to the analyst's input. This user interface provides an inherent anchoring bias to the Wall Street consensus, which prevents forecasts from being completely random numbers. Of course, this approach also results in consensus forecasts which can be fairly close to sell side numbers.

Secondly, estimates that appear to be erroneous are flagged as such in the database. The flagging uses a combination of automated error detection and human intervention. In our studies, we have removed flagged estimates. Approximately 2% of estimates in the system are flagged as potentially erroneous. The Estimize consensus is constructed by taking the most recent FQ1 estimate from each contributor, removing those estimates which are flagged, and calculating the equally weighted mean of the remaining estimates. In future studies, we will examine a Select Consensus which weights the cross section of estimates more intelligently based on the track record of the contributing analysts, the timeliness of their forecasts, and other characteristics.

For the analysis in this paper, and in other studies, we defined strict in sample and out of sample data sets, in order to develop more rigorous hypothesis testing and validation. Our in sample included all fiscal periods with report dates in 2011 and 2012, leaving reports in 2013 out of sample. We furthermore took a cross sectional in sample by selecting a random ¾ of names in each quarter, with the in sample and out of sample having the same representation by sector and by amount of Estimize coverage.

For all studies herein, we require there to be 3 outstanding, non-flagged estimates on a given name at the time of analysis. We also define an investible universe by restricting to U.S> names with market capitalizations of at least USD $100mm, an average daily trading volume over the prior month in excess of USD $1mm, and a closing price of at least $4. The universe is updated monthly throughout the sample and includes between 2,100 and 2,500 names on any given date in the sample, though many of these names will not have sufficient coverage on the Estimize system.

For our return-based studies, we follow standard industry practice by calculating daily residual returns within our universe, residualizing total return to the following common risk factors: industry, dividend yield, volatility, momentum, size, value, growth, and leverage. In unreported results, our findings are qualitatively the same if we simply take excess returns over the S&P 500 index. All of our results shown here are gross of both implicit and explicit transaction costs.

## ACCURACY OF ESTIMIZE FORECASTS

We first want to know whether the Estimize community is more accurate than Wall Street. To do this, we simply count the percentage of time the Estimize consensus is closer to actual earnings than is Wall Street as of the report date, restricting to cases where the two are different and rounding EPS estimates to the nearest penny and Revenue estimates to the nearest USD $1mm.

| | EPS | | | | Revenue | | | |
|---|---|---|---|---|---|---|---|---|
| | n | % more accurate | Estimize error | Wall Street error | n | % more accurate | Estimize error | Wall Street error |
| >= 1 analyst | 4890 | 51% | 17.5% | 17.4% | 4806 | 45% | 5.0% | 4.9% |
| >= 3 analysts | 2618 | 58% | 14.1% | 14.6% | 2645 | 49% | 3.8% | 3.7% |
| >= 10 analysts | 821 | 63% | 12.1% | 12.9% | 848 | 53% | 2.7% | 2.8% |
| >= 20 analysts | 258 | 65% | 13.5% | 14.7% | 264 | 54% | 2.6% | 2.8% |

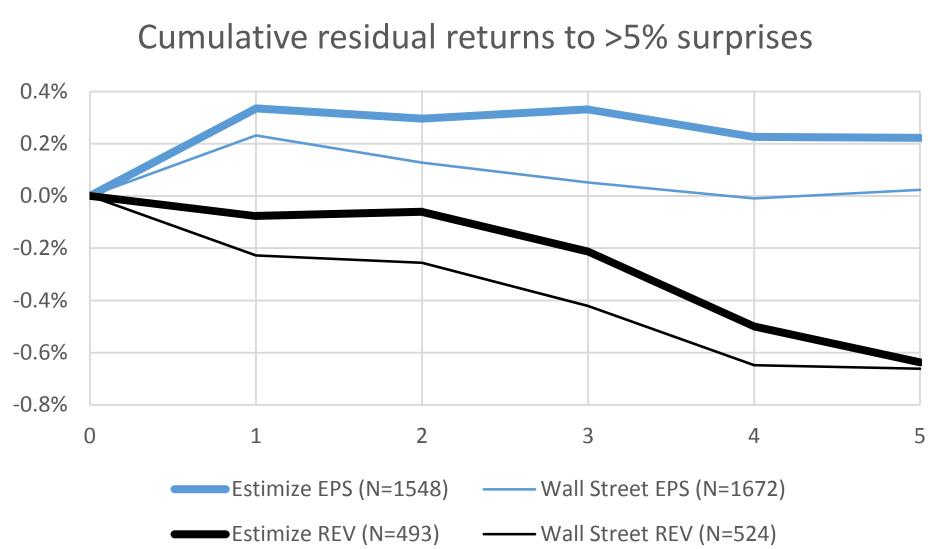
These results demonstrate that the Estimize EPS consensus is consistently more accurate than the comparable Wall Street number, and that the relative accuracy increases as the number of contributors increases. If we require at least 3 Estimize contributors, we also see that the average error of the Estimize consensus is lower than the Wall Street error. In the case of Revenue, we don't see an improvement until we restrict to heavily covered names (10 or more estimates). We also note that on average revenue errors are quite low, indicating that Wall Street analysts are already quite accurate with Revenue forecasts and that there is little room for improvement.

## POST EARNINGS ANNOUNCEMENT DRIFT

There is a long literature detailing the post earnings announcement, or unexpected earnings, drift effect [Bernard and Thomas, 1990]. Companies which beat expectations, as measured by Wall Street forecasts, have historically outperformed. However in recent years the effect has been greatly diminished, or at least compressed to a short enough time frame that investors without low-latency trading systems are unable to exploit the effect. In part this is likely due to the market's knowledge of the effect, as systematic traders have built algorithms to capture whatever alpha remains [Johnson and Schwartz, 2000].

Such algorithms depend on using data from Wall Street firms to measure the market's expectations. However, true market expectations may not be reflected in the sell side number. To examine whether the Estimize consensus provides an alternative, and possible more accurate, estimate of expectations, we measure the market response to earnings and revenue surprises as benchmarked separately against Wall Street and against Estimize.

For this study, the Day 1 return is calculated from the open to the close on the announcement date for names reporting earnings before trading opens, and from the next day's open to the next day's close for names reporting earnings after trading closes. Subsequent days' returns are measured close-to-close. In the chart and table below, we present the cumulative residual returns to surprises, as defined by actual earnings exceeding or falling short of expectations by 5% or more. The returns are signed (i.e., are reversed for negative surprises) and we take an equally weighted mean. We also present the Information Coefficients (IC), the rank correlation between the surprise magnitude and subsequent residual returns. In IC calculations we do not impose a minimum surprise magnitude constraint.



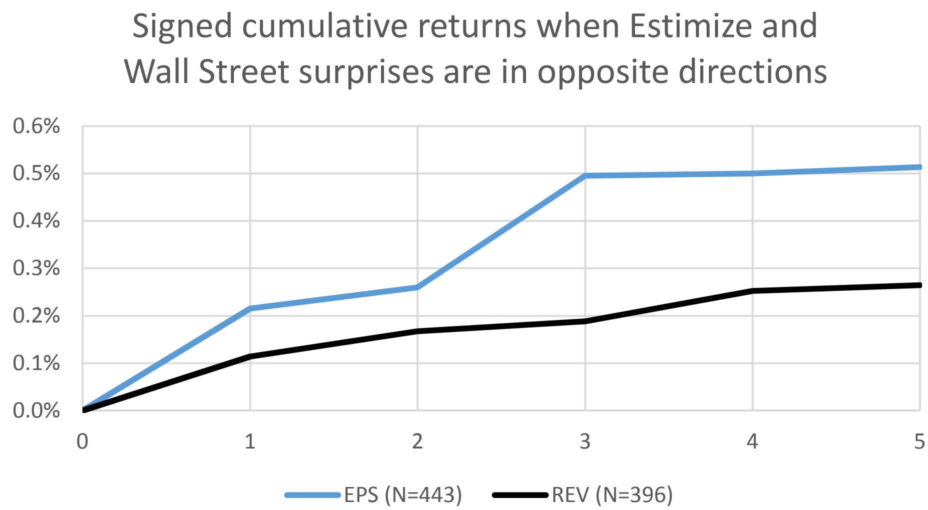
**Surprise return characteristics**

| | Nov 2011 to Jul 2013 | Estimize | | | | Wall Street | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | N | 1 day | 2 day | 5 day | N | 1 day | 2 day | 5 day |
| EPS | IC | 2771 | 0.048 | 0.044 | 0.026 | 2771 | 0.018 | 0.004 | (0.005) |
| Mean return | All surprises | 2737 | 0.25% | 0.20% | 0.21% | 2606 | 0.15% | 0.06% | -0.02% |
| | > 1% surprises | 2449 | 0.25% | 0.19% | 0.19% | 2458 | 0.15% | 0.05% | -0.01% |
| | > 5% surprises | 1548 | 0.33% | 0.30% | 0.22% | 1672 | 0.23% | 0.13% | 0.02% |
| | > 10% surprises | 1010 | 0.41% | 0.40% | 0.39% | 1120 | 0.22% | 0.11% | -0.12% |
| Revenue | IC | 2771 | (0.004) | 0.003 | 0.002 | 2767 | (0.000) | 0.014 | (0.005) |
| Mean return | All surprises | 2769 | 0.02% | 0.06% | 0.05% | 2766 | 0.00% | 0.02% | -0.02% |
| | > 1% surprises | 1933 | 0.02% | 0.09% | 0.09% | 1919 | 0.01% | 0.05% | -0.02% |
| | > 5% surprises | 493 | -0.08% | -0.06% | -0.64% | 524 | -0.23% | -0.26% | -0.66% |
| | > 10% surprises | 167 | -0.03% | 0.01% | -0.77% | 165 | -0.22% | -0.20% | -1.06% |

Using Wall Street as the benchmark, the post-EPS announcement effect lasts only for one day and is fairly small for such an infrequent event (15bps). With Estimize as the benchmark, the returns persist for several days without reverting, and are of a significantly greater magnitude. We also see larger drifts for larger surprise magnitudes when benchmarking against Estimize; the first day return is 25bps for 1% surprises but 41bps for 10% surprises.

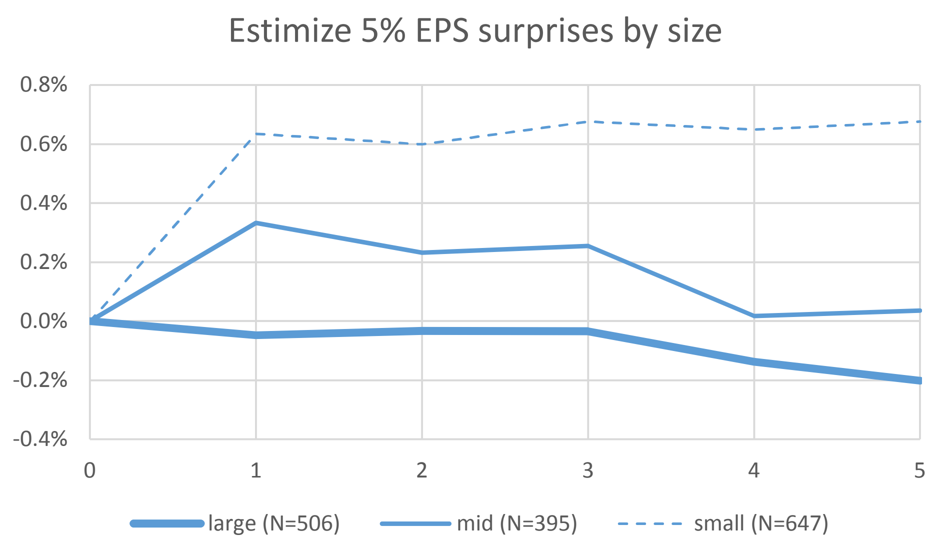
The Revenue surprise effect is not in evidence using either benchmark; in fact, it is even perverse if Wall Street is used as the benchmark. For most but not all names, Revenue is not as prominent a figure as EPS; furthermore, we find much less deviation between Estimize and Wall Street consensus numbers, and between either forecast and actual results when looking at Revenue, as evidenced by the sample sizes dropping off quickly as we look at larger-magnitude Revenue surprises. As a result, for the rest of the paper we focus on EPS estimates.

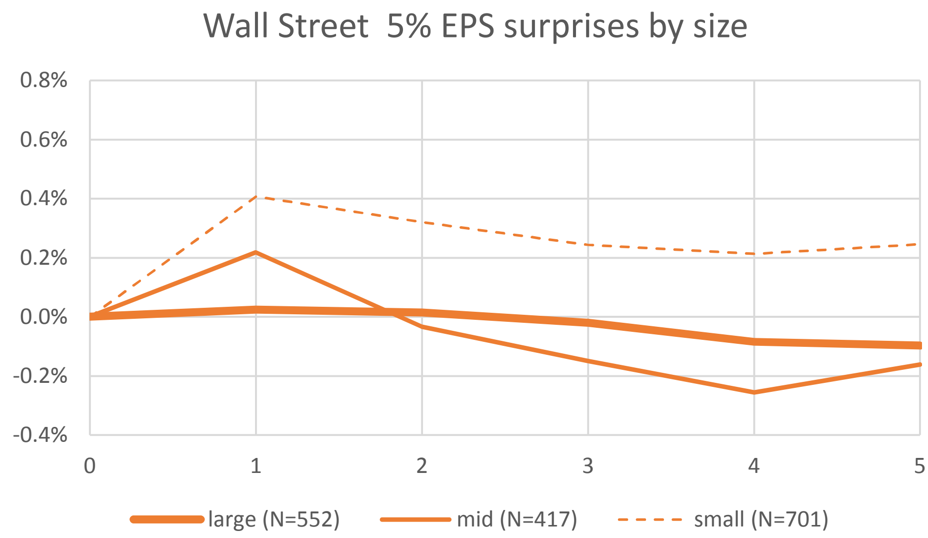
As an additional check, we examine the cumulative signed residual returns when there is a discrepancy between the surprises. In other words, we go long if actual earnings exceeds the Estimize consensus but falls short of the Wall Street consensus, and we go short if the opposite is true. While this study has a smaller sample size, it does confirm our hypothesis that, considered together, the Estimize consensus is a more accurate measure of market expectations.




Robustness checks

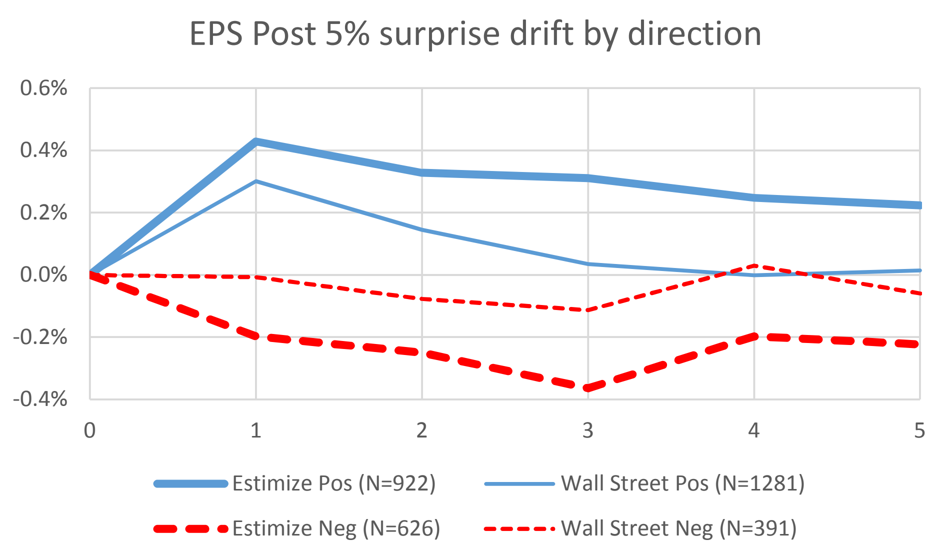
Next, we look at the signed surprise effect partitioned by size for each of the two benchmarks. Here, and in subsequent studies, we consider large caps to be the 500 largest U.S. names as measured by market cap that month; mid caps to be the next 500 largest names; and small caps to be the remainder of the universe.






In both charts, we see that the post earnings announcement effect is nonexistent among large caps. These names are subject to significantly more attention and news, and are on average far more liquid. Therefore we should expect the market to adjust to new earnings information much more efficiently. Among mid and small cap names, we see larger drifts when benchmarked against the Estimize consensus than among the Wall Street consensus. Among small cap names, the Estimize drift appears to last beyond the first trading day after the announcement.

Most earnings surprises are positive; companies are quite adept at managing expectations to create earnings surprises. 77% of earnings surprises in our sample, as measured relative to Wall Street are positive. While the Estimize community also exhibits such a bias, only 60% of Estimize-relative surprises are positive. We examine whether the results above are driven entirely by positive surprises by breaking out positive and negative cases. Note that in the following chart we do not sign the returns, so we expect a positive drift for positive surprises and a negative drift for negative surprises.



The results appear to be broadly similar for both directions, with longer and larger drifts when benchmarking against Estimize. In the case of Wall Street, there does not appear to be a significant drift after negative surprises, though the sample is relatively small given the preponderance of positive Wall Street surprises.

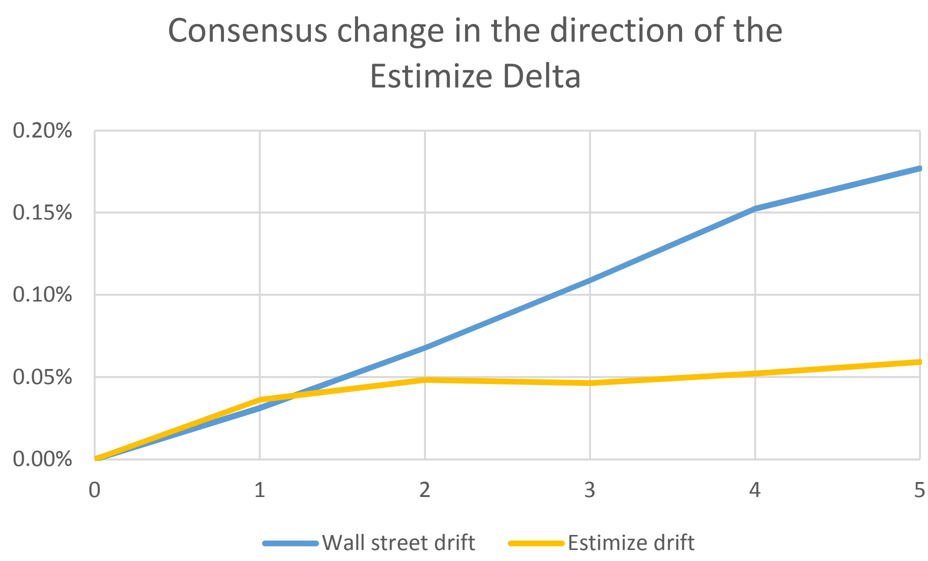
In unreported results, we also examined whether the findings were robust to sector. We found that the returns to a post earnings announcement drift strategy were stronger when benchmarked against Estimize than when benchmarked against Wall Street for every sector with at least 150 Estimize surprises.

## PRE EARNINGS EFFECTS: THE ESTIMIZE DELTA

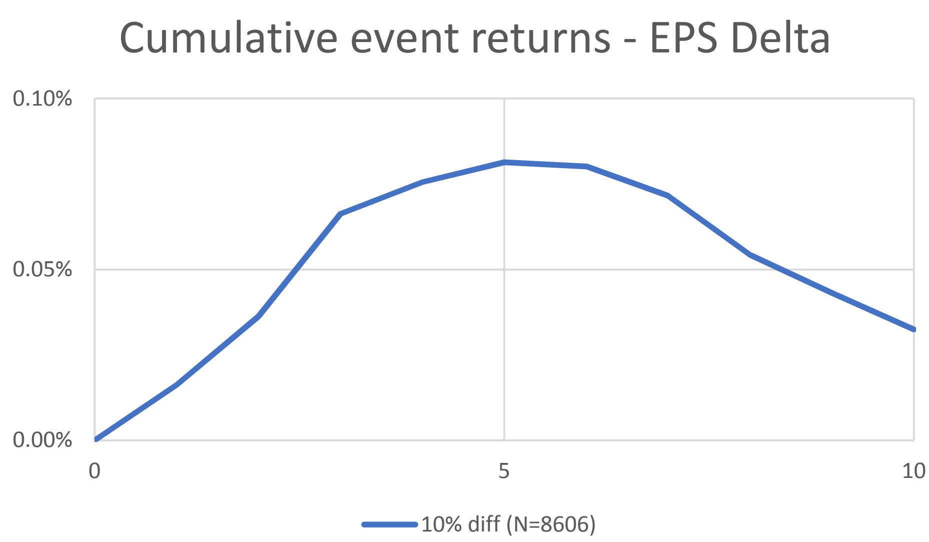
Many institutional investors track changes in sell side earnings expectations as a key input into their investment process, and the abnormal returns to such a strategy have long been documented in the academic literature [Givoly and Lakonishok, 1979]. Over the course of the quarter leading up to a report date, Wall Street analysts can update their forecasts several times to incorporate news, changes in outlook or company structure, macroeconomic conditions, or other factors. Similarly, contributors to the Estimize community can also alter their forecasts to incorporate new information. It is possible that the Estimize consensus may be more nimble because of the lack of institutional constraints on Estimize contributors. To examine this, we define the **Estimize Delta** as the percent deviation between the Estimize EPS estimate and the Wall Street EPS estimate in the days leading up to the earnings report date. To account for unusually high deltas due to small divisors, we impose a USD $0.10 minimum divisor in our calculation, so that if the Wall Street consensus is $0.02 and the Estimize is $0.04, we calculate the Delta as 20% rather than 100%.

When a Delta appears, we want to measure whether subsequent revisions occur in the same direction. If the Estimize consensus is more timely than the Wall Street consensus, we would expect forecast revisions in the latter following a significant Delta. It is also possible that serial correlation in changes in the Estimize consensus – due to herding effects, gradual information dissemination, differential timeliness across Estimize analysts, or for that matter incremental estimate revisions for a single analyst who is reluctant to make a single large revision – mean that when such a Delta appears, subsequent revisions among the Estimize community will also be in the same direction.

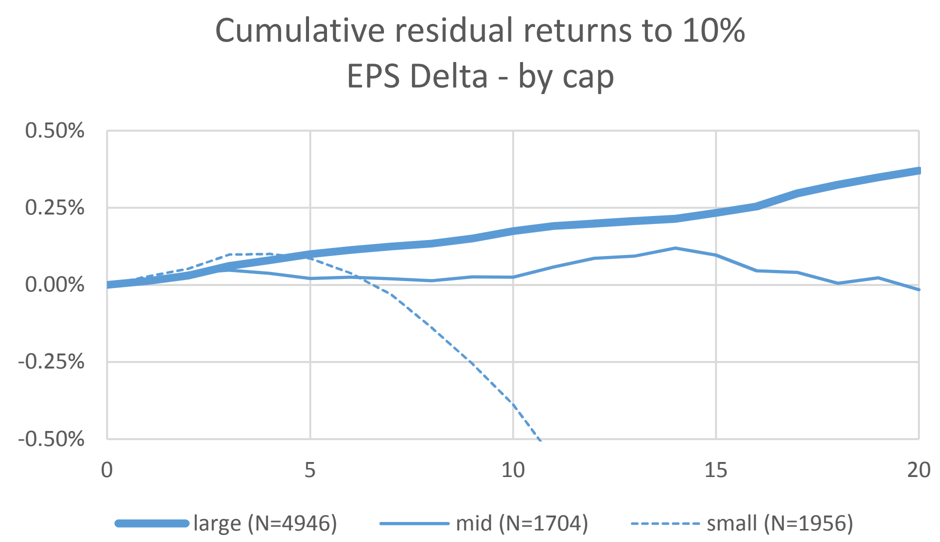
A look at future consensus changes in cases where the Delta is 5% or more, as a function of number of trading days after the Delta, confirms that the Estimize Delta predicts future changes in the Wall Street consensus, but that serial correlation within the Estimize community is more limited.




Institutional investors, however, trade on numbers that they receive from the sell side.  As such, the Estimize Delta should provide an early indication of such trading.  To measure this, we track the cumulative daily residual return as a function of number of trading days after a 10% or larger Delta signal is generated, close out signals the day before the company reports, since both expectations become obsolete once earnings are known.




We see a drift of between 5 and 10 basis points over the week following such a signal.  While this is obviously quite modest, a breakdown by capitalization provides some more interesting insights.




The cumulative residual returns are much stronger and longer lasting among large caps. Indeed, the returns last up to a month (though in some cases earnings are reported before the month is up) and have magnitudes exceeding 25bps. Among small and mid cap names, there appears to be little alpha beyond the first week. For these smaller names, we should note that the sample sizes drop off very quickly as we move to the right in the chart, since most estimates on these thinly covered names are made immediately before the earnings report date.

The capitalization difference is precisely the opposite of what we usually see with well-known investment anomalies such as sell side earnings estimate revisions strategies, indicating that the information in the Estimize has not been arbitraged away by market participants. This differential effect may also reflect the existence of more, and better quality, data among large caps; note the sample sizes. Finally, to the extent that the abnormal returns are the result of future revisions among sell side analysts, the heavier coverage among larger stocks may make such a forecast more reliable. The stronger returns among large caps is good news in that transaction costs and market impact should be much less severe for an investor who trades on the Estimize Delta among more liquid names.

As an investment factor, then, the Estimize Delta looks quite attractive among large cap U.S. names. The returns in our studies are residualized to common risk factors, so we know that the returns are not explained by those factors. However, it is still possible that the Delta has exposures to those risk factors, or that it is redundant with a sell side-derived earnings revision alpha factor. Indeed, such factors may be profitable precisely because they take advantage of the serial correlation of EPS revisions to predict future changes in sell side forecasts.

A cross sectional examination of the correlation of the Estimize EPS Delta with changes in Wall Street forecasts, and with our risk factors, reveals no significant exposures.

| Correlation of EPS Delta to other factors | | |
|---|---|---|
| Change in Wall Street consensus: | 1 day | 0.03 |
| | 5 days | (0.01) |
| | 10 days | (0.07) |
| | | |
| Risk factors: | yield | 0.08 |
| | volatility | (0.07) |
| | momentum | 0.01 |
| | size | 0.02 |
| | value | 0.08 |
| | growth | (0.05) |
| | leverage | (0.00) |

So it appears that the Estimize Delta is indeed an orthogonal and potentially profitable investment signal that is uncorrelated with existing sources of alpha or risk.

## FUTURE RESEARCH

Throughout this paper we have focused on an equally weighted consensus estimate. Estimize has recently released the Select Consensus, a re-weighted forecast which takes into account the sector-level track record of an analyst, the difficulty of estimation of the names he or she covers, the optimistic or pessimistic bias of the analyst, the amount of coverage the of the analyst (both over time and cross sectionally), and the recency of the estimate. Early research shows that track record is a persistent characteristic of contributors: contributors who were accurate in the past are likely to remain accurate in the future. It also seems that the Select Consensus improves on the accuracy of a simple mean Estimize estimate. As such, the alpha factors shown in this paper may be enhanced by using such a re-weighted estimate.

A more accurate earnings estimate may have other uses as well, for example in valuation, margin estimation, or in aggregates to the industry or sector level.

Estimize will eventually support other measures beyond EPS and revenue estimates, for example EBITDA, margins, or macroeconomic forecasts. All of these have the potential for providing rich new data for institutional investors.

Finally, the Estimize database is growing over time, both in terms of instruments covered and number of contributors. At some point in the future non-U.S. names will be covered as well. And a longer history will allow us to do more robust studies over a variety of investment cycles, and to measure the track record of contributors more accurately.

## CONCLUSIONS

The above results indicate that the Estimize consensus is an accurate and leading measure of market expectations. As a result we can generate trading strategies both before and after earnings reports to capture abnormal returns using Estimize data. The pre-earnings announcement strategy using the Estimize

EPS Delta is particularly strong and long lasting among liquid, large cap names, indicating a potentially profitable signal that is furthermore orthogonal to existing measures of risk and return.

These results represent a simple starting point for research on crowdsourced financial forecasts. There is much more that can be done with the existing data, and other forecasts will become available over time as well which will leverage the wisdom of crowds to generate actionable alpha.

## REFERENCES

Avery, Christopher, Judity A. Chevalier, and Richard Zeckhauser. "The 'CAPS' Prediction System and Stock Market Returns". NBER Working Paper No. 17298, 2011

Bernard, Victor L., and Jacob K. Thomas. "Evidence that stock prices do not fully reflect the implications of current earnings for future earnings." Journal of Accounting and Economics 13(4), 1990

Boni, Leslie, and Kent Womack. "Solving the Sell-Side Research Problem: Insight from Buy-Side Professionals." Dartmouth College working paper, 8 August 2002

Givoly, Dan, and Josef Lakonishok, 1979. "The information content of financial analysts' forecasts of earnings: Some evidence on semi-strong inefficiency." Journal of Accounting and Economics 1, 1979

Hong, Harrison, Jeffrey D. Kubik and Amit Solomon. "Security Analysts' Career Concerns and the Herding of Earnings Forecasts." RAND Journal of Economics 31, 2000

Johnson, W. Bruce, and William C. Schwartz Jr. "Evidence That Capital Markets Learn from Academic Research: Earnings Surprises and the Persistence of Post-Earnings Announcement Drift." University of Iowa working paper, 2000.

Michaely, Roni, and Kent L. Womack. "Conflict of Interest and the Credibility of Underwriter Analyst Recommendations." Review of Financial Studies, 1999, vol. 12, issue 4

Trueman, Brett. "Analyst Forecasts and Herding Behavior." The Review of Financial Studies Vol. 7, No. 1, Spring 1994

Wolfers, Justin, and Eric Zitzewitz. "Prediction Markets in Theory and Practice." NBER Working Paper No. 12083, March 2006