

An EPS Beats Predictor Featuring High-Accuracy, Daily Updates, and Explainability

October 2023

Weihong Zhang, PhD
CEO
Cmind Inc
wzhang@cmind-ai.com

Executive Summary

One of the most important and closely watched numbers on Wall Street is a firm’s quarterly earnings per share (‘EPS’) and how it compares with analysts’ consensus EPS estimates. Cmind Inc developed an EPS beats predictor that predicts whether a company beats or misses its earnings target one quarter ahead of its earnings release. The predictor starts predicting right after a company files its 10Q of a reported quarter, and updates its predictions daily until the next quarter’s earnings release by taking account of newly available information from the market.

The predictor combines six heterogenous data modalities with different frequencies including NLP signals extracted from management commentary. It features high accuracy (prediction cohort-based accuracy proposed to provide more granular analysis), high-frequency update (daily updates allowing investors a long lead time prior to decision-making), and explainability (target proprietary variables available to explain the prediction outcome).

1. Stock Movement vs EPS Beats

One of the most important and closely watched numbers on Wall Street is a firm’s quarterly earnings per share (‘EPS’) and how it compares with analysts’ consensus EPS estimates. Firms whose actual EPS beat (miss) the consensus forecast are said to experience a positive (negative) earnings surprise. EPS beat/miss is often followed by significant subsequent stock price movements for as long as two months, a phenomenon termed “Post Earnings Announcement Drift” first documented by Ball and Brown (1968). This is great news for traders and portfolio managers alike as accurate predictions about firms’ EPS beats/misses can be a valuable signal for executing profitable traders. The earlier the prediction, the more lead time that investors can make decisions and more profit opportunity that investors might be able to capitalize.

Cmind Inc performed an analysis in how the EPS beats/misses influence the stock movement in a short-term basis. We selected 4-year data from 2019Q2 to 2023Q1. We chose to purchase the beats/misses stocks at the closing prices at the close of the earnings release days and hold for 2 days. The results are as shown in Table 1. Varying the holding period from 1 to days leads to similar investment results.

Table 1: Average 2-day investment return of US-listed companies vs Beats/Misses

Companies	Average Gain/Loss
Beats	+1.53%
Misses	-2.58%

2. Current Practice of Estimating EPS Beats

To predict earnings, most analysts build financial models that estimate future revenues and costs. For revenue forecast, analysts estimate sales volume growth, estimate the prices companies can charge for the products and other factors. On the cost side, analysts look at expected changes in the costs of running the business. Costs include wages, materials used in production, marketing and sales costs, interest on loans, etc. As of December 2021, of the approximately 9,000 companies that traded on U.S. markets, over 50%

are neglected by sell-side analysts.¹ This is not surprising given the costs involved in providing such coverage.²

To create useful financial models of any corporation, analysts must process volumes of data from many disparate sources, both qualitative and quantitative. In a commonly used method, the top-down approach, analysts start with macroeconomic growth rates, currencies, interest rates, or other factors that impact corporate earnings and growth. To further understand the dynamics of the individual companies they cover, analysts will also speak to customers, suppliers, and competitors to get a sense of the strength of demand for the company's product as well as any competitive advantage the company has. In addition to their own independent research, analysts often supplement their data with companies' forward-looking guidance, commentary from management at earnings calls and industry conferences as well as news coverage and press releases between earnings calls and build that into their valuation models. Overall, the current practice of EPS forecasting is both time consuming and labor-intensive, resulting in most analysts covering fewer than 25 stocks.

Since the invention of the Internet, humans have begun a transformational revolution of shifting everything online. This shift, coupled with the explosion in the availability of real-time data available from traditional sources like news outlets and social media such as Twitter, WallStreetBets, etc., poses new challenges for asset managers because of their high frequency, enormous quantity, and high signal-to-noise ratio. While standard financial information such as 10Q's and 10K's are easily accessible, it is especially difficult to combine this numerical information with unstructured text based alternative data sources, traditional news and management-issued guidance in an accurate manner and to extract useful information from it. Common practices that are currently in place rely on expensive manual labor to read through thousands of threads to extract embedded useful information, which is rather inefficient and unsustainable in the long run.

Despite best efforts from many large incumbent firms who analyze earnings (e.g., Starmine of Refinitiv) and small firms (e.g., Estimize) alike, current accuracy of EPS prediction is only slightly better than a coin toss in select periods. Starmine's uniqueness is to overweight the earnings predictions from high-quality analysts, but the analysts usually exhibit mean reversion in their forecast accuracy. Estimize's approach for EPS prediction is to crowdsource earnings and macroeconomic estimates from many contributors across the globe. However, this approach contains a lot of noise and is subject to random errors. Regardless of the approach, the market has a great demand for an EPS prediction tool which covers the entire market that is automatic, real-time, and has greater accuracy.

3. Data Science to the Rescue

Due to the significant amount of time and effort required of analysts to analyze companies' EPS in depth, most analysts specialize in only a handful of firms in a single industry, leading to many potential missed opportunities. In particular, while large firms such as Apple Inc. and Amazon.com Inc. attract significant analyst coverage due to greater coverage by sell side investment banks, numerous small firms suffer from scant or even no analyst coverage. With the increasing availability of data both from public sources and from management commentaries, artificial intelligence can come to the rescue by automating the collection of vast quantities of data and processing it into usable signals for EPS beat/miss prediction. This would significantly improve analysts' ability to cover a wide range of industries and companies at a very low cost for the firm. However, achieving a high accuracy in out of sample predictions for EPS beat/miss is a harder task than it appears.

¹ <https://www.zacks.com/education/articles.php?id=57>

² <https://www.cnb.com/2017/10/06/why-dozens-of-public-companies-are-ignored-by-analysts.html>

Given that the current practice largely relies on developing highly customized valuation and prediction models on a handful of companies that they have expertise or knowledge about, the approach fundamentally constrains analysts' scope of coverage.³ In addition, analyst coverage is disproportionately tilted towards large cap stocks that are household names, leaving many medium- to small-size stocks uncovered or covered by only a couple of analysts. This imbalance in analyst coverage results in reduced information production, which leads to market inefficiency and mispricing.

One alternative approach to security analysis is to cover many, preferably the universe of publicly traded securities, with automated processes leveraging frontier advances in Data Science and Artificial Intelligence. In contrast to the aforementioned status quo in securities analysis, this alternative approach can generate prediction models that are of high-accuracy, frequency, and consistency enabling relatively scarce and expensive analysts to focus their limited attention on those securities which are highlighted by the data science signals well in advance of the next release.

The holy grail in securities analysis should be an organic mixture of the two different approaches, striking a balance between highly customized industry-specific modeling and generalized accurate prediction for a general firm. That is, the latter automated approach can significantly enhance and empower analysts' current practices by improving their efficiency and scope of analyses.

4. Cmind's Methodology and Advantages

Cmind's methodology to EPS prediction is deeply rooted in modern data science, specifically by leveraging the frontier algorithms in Natural Language Processing (NLP) and Machine Learning (ML). NLP refers to a branch of Artificial Intelligence aimed at giving computers the ability to understand words in the same way that humans do. Importantly, NLP technology has transitioned from simple counting of keywords to extracting the meaning with the context considered, which is often referred to as semantic NLP. Machine Learning (ML) broadly refers to the study of computer algorithms that can improve automatically through domain expertise and data. Specific to Cmind, ML serves as the backbone technology to form predictive models using optimized algorithms after the ingestion and annotation of databases. Cmind also develops an in-house set of proprietary annotations that are accurate and conform to both domain knowledge and expert opinions. The data sources employed range from publicly available quarterly financial data and macroeconomic variables to company-specific earnings call transcripts, management commentary, and corporate governance variables.

Figure 1 shows the data sources and methodology of the Cmind EPS Beats/Misses Predictor. It leverages over 10 years of EPS beat/miss history, 40 years of company quarterly financials, 4 years of earnings transcripts, 20 years of corporate governance variables, 2 years of news, and 40 years of macroeconomic variables. These signals have been trained by our powerful Machine Learning algorithm optimized for time series modeling and NLP. These techniques together enable the Cmind EPS Predictor to achieve a market leading prediction accuracy in select industries.

The Cmind EPS predictor accurately predicts US company earnings beats and misses for over 4,500 publicly traded companies. Importantly, it can do this a full quarter (i.e., 90 days) ahead of the next release using frontier machine learning to extract signals from accounting statements, earnings transcripts, economic variables, and consensus EPS. The Predictor has the following advantages.

- **High accuracy.** Built upon 6 heterogeneous data modalities and 150+ target input variables, the predictor can generate probabilities that highly correlate to the beats or misses outcomes.

³ Most analysts are found to cover fewer than 15 individual stocks.

Furthermore, by categorizing the probabilities into different **prediction cohorts**, the prediction accuracies can be further improved. See next sector for more details.

- **Early prediction and high-frequency update:** Predictions are generated very early for a company in that they are generated soon after a company files its 10Q of a report quarter. The predictions are updated daily all the way until its next earnings release. This gives investors a long lead time to observe the trends prior to making their investment decisions. As an illustration, Figure 2 shows that Airbnb, Inc. (\$ABNB) is Very Likely to Beat its consensus EPS forecast in its Q3:23 earnings release scheduled on November 1st, 2023. The plot was drawn as of October 27th. Historically, Cmind has correctly predicted ABNB Beats with 90% accuracy. Cmind started predicting its Q3 EPS beat probability from early August. Over the period of almost 3 months, its predictions have been consistently pointing to a beat with high probabilities.
- **Insights and Explainability:** The predictor provides a rich set of explanatory variables that represent a company’s financial reporting risks, the linguistic patterns of its executives and analysts, and comparisons against its competitors in a time series manner. These proprietary variables are available to explain the prediction outcome.

Figure 1: Cmind Methodology to EPS Beats Prediction

Cmind EPS Predictor Construction

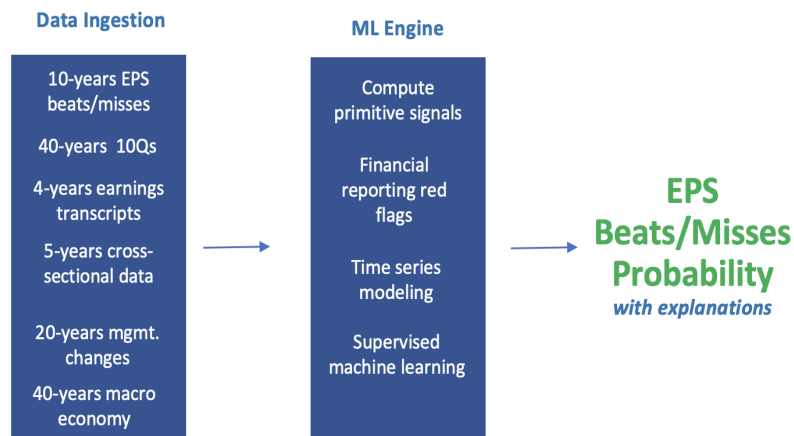
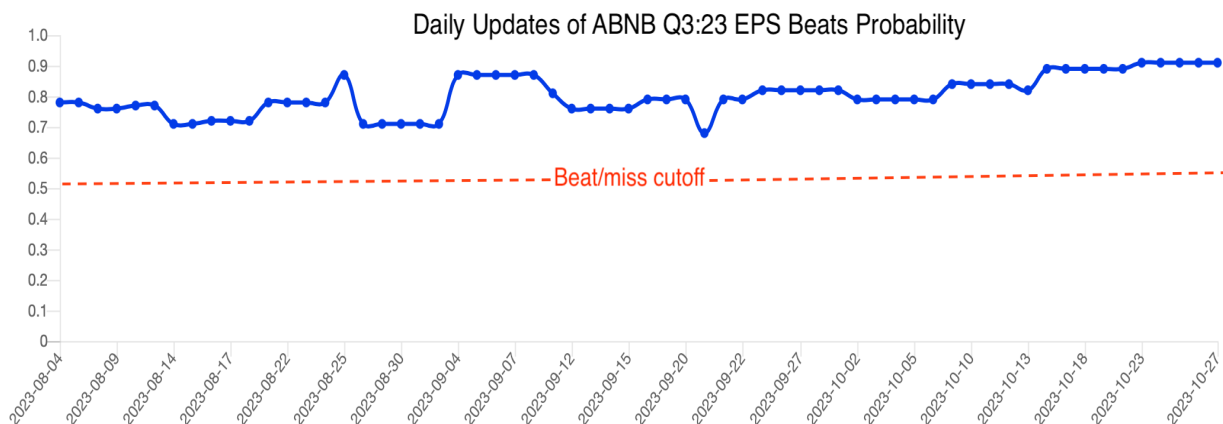


Figure 2: Cmind EPS Beats Daily Predictions



5. Accuracy vs Prediction Cohorts

The output of The Cmind EPS Predictor is a probability that falls into the range [0.0, 1.0]. To calculate the accuracies of a predictor, the numerical probability needs to be converted to predicted beats or misses categories. This step is called *discretization*. There are different ways of discretization.

- Oftentimes, it is to simply pick a cut-off threshold, say 0.5; the probability is interpreted as a beat if it is larger than the threshold, and interpreted as a miss otherwise. This discretization approach essentially leads to two *prediction cohorts* – *predicted beat* or *predicted miss*, depending on whether the magnitude of the probability falls into [0.0, 0.5] or [0.5, 1.0].
- The discretization can be more granular. For instance, we can divide the probability range [0, 1.0] into five equal intervals, namely [0, 0.2], [0.2, 0.4], [0.4, 0.6], [0.6, 0.8], and [0.8, 1.0]. For apparent reasons, we call them *very likely miss*, *likely miss*, *marginal*, *likely beat*, and *very likely beat prediction cohort*.

By dividing predictions to different cohorts, we can analyze the accuracy in a more granular level, which results in higher accuracy at certain cohorts than others. This leads to more reliable investment opportunities at high-accuracy prediction cohorts.

5.1 Accuracy in 2 Prediction Cohorts

Table 2 shows the accuracy results in 2 prediction cohorts across sectors, caps and indices. From 2020 Q4 onwards, the algorithm relatively consistently achieves greater than 65% accuracy for all 11 broadly defined industry sectors, and greater than 70% accuracy for select industries. The differential accuracy across various industries arises because of the common weighting scheme applied to achieve the greatest prediction accuracy for the overall sample. As Cmind accumulates a longer history of time-series NLP data and tunes the model further, the algorithm’s accuracy will continue to improve in the foreseeable future.

In different capitalizations, the Cmind EPS Predictor achieves an average accuracy of 77%, 71%, and 66% respectively for large, mid, and small-size companies over the last 3 three years. In addition, in different indices, the predictor achieves an average accuracy of 79% and 77% respectively for S&P 500 and Russell 1000 firms over the last 3 years.

Table 2: Accuracy in 2 Prediction Cohorts across sectors, caps, and indices

	2019Q3	2019Q4	2020Q1	2020Q2	2020Q3	2020Q4	2021Q1	2021Q2	2021Q3	2021Q4	2022Q1	2022Q2	Average
Capitalizations													
Large	75%	73%	71%	74%	81%	79%	84%	84%	78%	77%	75%	73%	77%
Mid	69%	66%	67%	66%	80%	76%	77%	79%	70%	68%	69%	69%	71%
Small	64%	63%	64%	58%	75%	73%	72%	71%	63%	63%	62%	64%	66%
Sectors													
Communication Services	66%	76%	71%	67%	73%	72%	75%	72%	64%	66%	68%	69%	70%
Consumer Discretionary	63%	60%	67%	61%	80%	74%	74%	77%	64%	71%	63%	64%	68%
Consumer Services	76%	79%	65%	73%	81%	82%	73%	73%	68%	66%	68%	69%	73%
Energy	66%	62%	65%	57%	62%	66%	67%	69%	65%	60%	62%	69%	64%
Financials	63%	66%	61%	45%	82%	81%	83%	82%	72%	71%	68%	64%	70%
Health Care	66%	63%	66%	63%	72%	70%	66%	67%	62%	63%	62%	61%	65%
Industrials	68%	64%	69%	70%	83%	71%	76%	78%	65%	67%	69%	68%	71%
Information Technology	70%	71%	69%	70%	84%	81%	82%	81%	73%	69%	71%	72%	74%
Materials	72%	69%	65%	72%	79%	80%	77%	73%	72%	64%	72%	69%	72%
Real Estate	69%	54%	60%	55%	60%	69%	72%	72%	67%	61%	65%	70%	64%
Utilities	69%	60%	62%	67%	68%	67%	77%	75%	66%	73%	59%	65%	67%
Indices													
SP500	75%	74%	74%	74%	82%	81%	86%	87%	82%	80%	79%	76%	79%
Russell 1000	74%	71%	69%	72%	82%	80%	84%	84%	77%	77%	76%	72%	77%

5.2 Accuracy in 5 Prediction Cohorts

When it comes to the accuracy regarding 5 prediction cohorts, it has much more combinations of sectors, caps, and cohorts. For each sector, one can measure the accuracy at each of the 5 cohorts; similarly for each cap, one can measure the accuracy at each of the 5 cohorts.

Figure 3 shows the accuracy of the “very likely beat” cohort whose probabilities are in the range [0.8, 1.0]; from 2018Q2 to 2023Q3. Its average accuracy over all the quarters is 86%.

Figure 3: Accuracy in the “very likely beat” cohort for the large cap

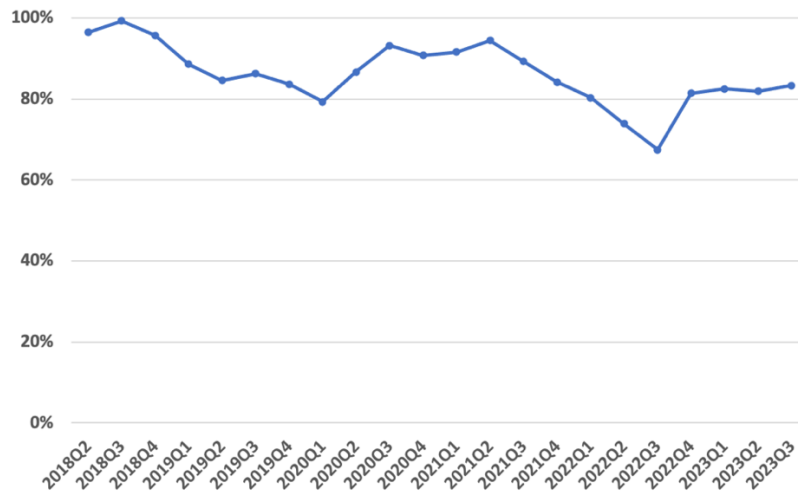
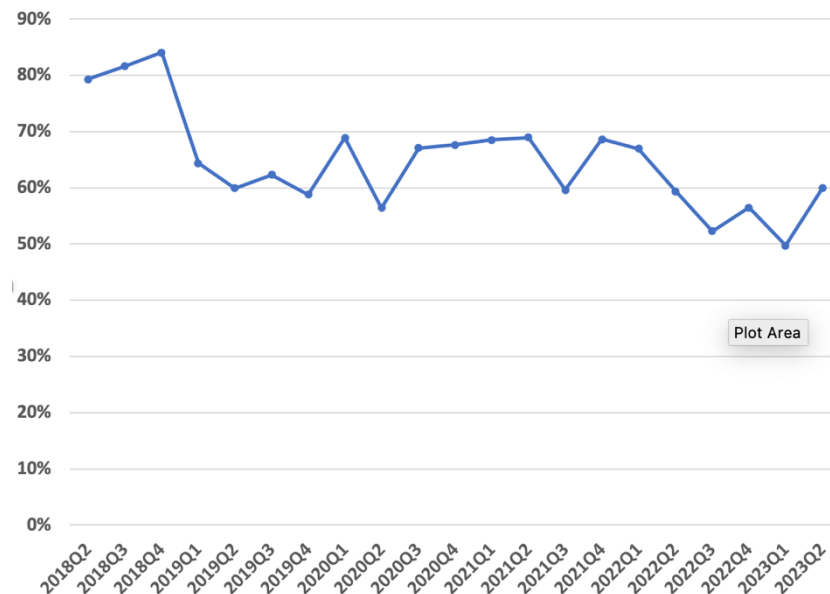


Figure 4 shows the accuracy of the “very likely miss” cohort whose probabilities are in the range [0, 0.2]; from 2018Q2 to 2023Q3. Its average accuracy over all the quarters is 65%.

Figure 4: Accuracy in the “very likely miss” cohort for the small cap



6. Implications for Traders and Portfolio Managers

With an overall accuracy rate that is consistently greater than 65% and higher accuracy in different prediction cohorts, the Cmind EPS predictor creates significant new opportunities for traders, portfolio managers, and analysts alike. In particular, it offers a cost-efficient way to deepen the understanding of existing stocks in one's portfolio and/or to gain useful insights on a new stock under consideration. For traders, the predictor will enable them to exploit a reliable signal around quarterly earnings announcements; for portfolio managers, Cmind EPS predictor will provide a straightforward way to evaluate allocations within their portfolio well in advance of major earnings releases.

Different from most products that are available on the market, the Cmind EPS predictor offers a beat/miss prediction a full quarter before the release of the actual quarterly results. While the model is continuously updated by incorporating new information between the latest quarterly earnings announcement and the next, the early provision of an accurate beat/miss signal for the upcoming quarter gives sell-side and buy-side clients alike a chance to either re-examine their existing portfolio or gain insights into new investment opportunities. Cmind EPS predictor empowers analysts, traders, and portfolio managers by providing an early useful signal on one of the most important corporate events, earnings releases.

Third-party independent backtesting shows that Cmind beats/misses prediction probabilities can be used to gain investment alpha by appropriately picking stocks of the large probability beats and the small probability misses), for instance, if one trades stocks with beats probability greater than 0.80 at MOO (Market On Open) and holds for 1 day, the Sharpe and Sortino ratio could be 2.36 and 3.21 respectively.

About Cmind Inc

Cmind Inc combines management commentary analytics, leading-edge accounting principles, and artificial intelligence to provide investors, accountants, and lawyers with unique and timely forecasts of company risks, events, and financial performance. Cmind enables its clients to make fast and accurate decisions about corporate valuations, credit, and sustainability.

For more information, please visit <https://www.cmind-ai.com> or contact info@cmind-ai.com.