

4. LEVERAGING AI/ALTERNATIVE DATA ANALYSIS IN SELL-SIDE RESEARCH: GOLDMAN SACHS

Contributors: Ingrid Tierens, PhD, CFA²²
Dan Duggan, PhD²³

	Asset Allocation	Equity	Debt	Hedge Funds
Americas				
Asia Pacific				
Europe, Middle East, and Africa				

Background

Goldman Sachs' Global Investment Research Data Strategy team works directly with both equity and macro research analysts around the globe on projects that require analytical and quantitative skill sets, with the research analysts providing the subject matter expertise and spelling out the investment use cases. Over the past two-plus years, the team (based in New York City and Bengaluru, India) have collaborated on nearly 200 published research analyses across various sectors and markets worldwide. Data sets include app analytics, social media, satellite imagery, government-sourced data, and newspaper articles. Combining these sources with a variety of advanced techniques, such as clustering or sentiment analysis, can create unique investment research insights. One example is "GS Aggregates Share Tracker: Geospatial Assessment Points to Pricing Upside," published by lead equity analyst Jerry Revich, CFA, on 17 January 2019, which leverages publicly available geospatial data to create local market share estimates and map it to local demographics.

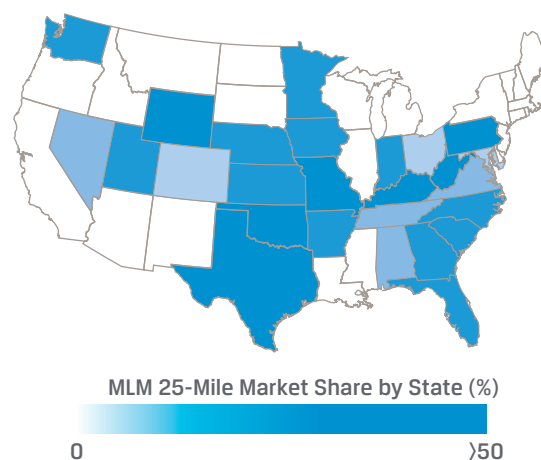
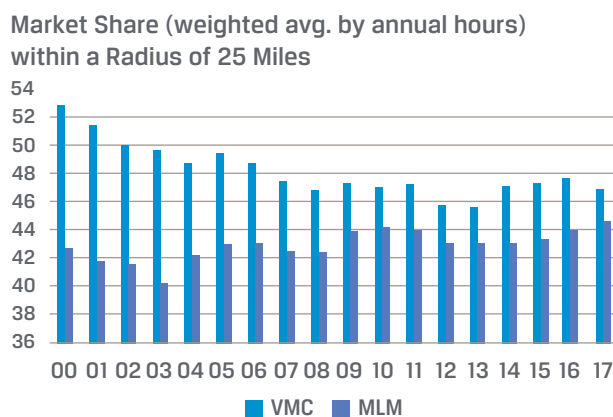
Analytical Approach

The GS Aggregates Share Tracker answers the following questions relevant to a subsector of the materials industry: (1) how to understand positioning in a hyperlocal industry, (2) how to represent market share for public companies in a largely private competitive landscape, and (3) how to inform investment professionals about the directional sense of quarterly company results with respect to aggregate volumes. In the aggregates industry (the mining of sand, gravel, and crushed rock for use in concrete for the construction industry), volume growth and product pricing margins are two of the most important key performance indicators. The markets themselves are

hyperlocal, meaning that market share calculated at the 25-mile range is more accurate than those at the state or even the county level. Identifying where a company's quarries are located as well as providing a quarry-level estimate of production volume deepen the analyst's understanding of how a company is positioned (e.g., in which markets a company has pricing power or material exposure). The addition of a new data source that meaningfully tracks company performance with geospatial information provides better insight into how these companies operate.

Figure 10, which reproduces Exhibit 21 and Exhibit 8 in Reviv's 2019 report, visualizes the added insight for Martin Marietta (MLM) by showing its 25-mile market share by state and comparing it with the market share of Vulcan Materials (VMC) over time.

FIGURE 10. MARTIN MARIETTA MARKET SHARE BY STATE AND COMPARISON WITH VULCAN MATERIALS MARKET SHARE



Sources: Mine Safety and Health Administration, company data, and Goldman Sachs Global Investment Research.

22 Managing Director, Global Investment Research, Goldman Sachs. This case is prepared based on our interview with her and materials submitted by Goldman Sachs.

23 Vice President, Global Investment Research, Goldman Sachs.

Estimating aggregates volumes in this space can inform a company's strategy: which markets are in focus, how potential M&A activity may change the company's positioning, and where pricing margins are most important. At its most basic level, the analysis is used to estimate organic growth on a quarterly basis and to provide directional input ahead of earnings. In addition, the analysis is used to understand market share shifts and pricing impact and is levered to estimate company-level exposure weighted by aggregate volumes to weather events, such as regional flooding and hurricanes.

AI/Alternative Data Analysis

The GS Aggregates Share Tracker combines both company data and publicly available quarry data as inputs to the model. It takes advantage of open-source geospatial libraries and leverages the quarry specific metadata to provide estimates of both aggregates production and location information. An iterative validation sequence allows corrections to the raw data, which reduces uncertainties associated with the derived locations. Entity mapping is performed through a combination of NLP and company-reported subsidiary data, in conjunction with the analysts' domain expertise. Market share calculations are location specific and thus are calculated for each of the roughly 9,000 US quarries at distances from 5 to 50 miles in 5-mile increments.

Team Structure and Development Process

To achieve our objective of offering buy-side clients better and more differentiated investment insights using alternative data and machine learning, we aim for the intersection of the following three functions: (1) tap into domain expertise, (2) access all relevant information, including both "traditional" sources and alternative data where appropriate, and (3) apply advanced analysis techniques to extract relevant insights.

Whether the project is situation specific or longer term, it is absolutely critical to first understand the investment research use case we are trying to solve. Otherwise, we could end up wasting a lot of time and effort. Communication between the research domain experts and the quants is key, and the more opportunities to understand the research process, the better.

Conversations with research analysts help us determine whether our value added as a team comes from embedding alternative data, providing data analysis expertise, or helping analysts systematize their insights into a data product. Sometimes we discover it is not the right project for us to be involved in.

The model creation process of the quant research team is fully integrated with the analyst team, and both teams are involved with building and validation from start to finish. For the GS Aggregates Share Tracker, our quant team worked with the covering analyst team led by Jerry Revich on every step of the process: iterating entity mapping results for each company and cross-checking company quarry location model mapping, historical aggregate production results, and specific validation sequences based on significant events on a per company basis. The initial models relied only on public company data, but as the model has grown in complexity, it has expanded to cover the entire 9,000+ US quarry industry, creating tremendous scope expansion for the analyst team.

Key Takeaways

Some advice based on our experience:

- Don't underestimate the potential of more advanced techniques and approaches for making investment professionals more efficient and thus freeing them up to spend more time on creating better alpha insights!
- Alternative data does not necessarily translate into an advantage for systematic managers. In fact, we think that more niche, sector-specific data sets lend themselves much better to a fundamental analyst or portfolio manager who digs deeply into specific assets as opposed to an analyst or portfolio manager who covers a much broader range of assets.
- In our experience, a single data set or a single methodology has never been the sole driver of a research product. For us, alternative data adds to the mosaic, but it is not a goal in and of itself.
- Good research analysts are often aware of data that is relevant to the assets they cover, but they may not know how to best access and/or analyze the data.
- Leveraging alternative data doesn't necessarily mean breaking the bank. There is quite a bit of publicly available data that can be additive when you properly combine it with other data sources you may already be looking into.