Managers of Geller, Bing, and Buffett (GBB) Investors:

I was hired to assess how contemporary analytics might be incorporated into the way GBB Investors values various assets, especially stocks in companies. Your company has largely relied on more traditional forms of asset valuation to determine when stocks are under- and overvalued. While this has served your company well in the past, smaller and more agile firms are starting to effectively leverage contemporary analytics and automated trading to significantly out-perform their competitors. I was hired to assess whether these tools could be incorporated into your asset valuation practices without losing what has made your firm so successful to date. I believe I have developed a solution that accomplishes these goals, and that implementing it will drastically improve GBB Investor's ability to identify lucrative investment opportunities.

Along with a brief conclusion and this introduction, I divide this report into three substantive sections. In the first, I summarize pertinent aspects of your current asset valuation process—which I uncovered during a sixth month observation period, throughout which I carefully followed everyday operations of GBB Investors. Importantly, I discovered practices that I believe lead to regular and substantially suboptimal investment decisions at your company. In the second section, I propose to replace these suboptimal practices with ones incorporating contemporary analytics, using state-of-the-art technology to make evidence-based projections to be used in asset valuation. In other words, I believe we can leverage AI to help you and your employees identify over- and undervalued assets. Then, in the third section, I acknowledge the potential drawbacks and shortcomings of my proposal.

SECTION I: Current Practices

The typical asset valuation performed at GBB Investors relies on **discounted cash flow analysis (DCFA)**. This model calculates the fair market value of an asset (e.g., a stock in a company) based on its current value and its future value, where the latter is discounted to account for the **time value of money**. GBB Investors has been a long-time leader in its field by effectively carrying out and leveraging DCFAs. As you surely know, one needs to estimate the following quantities (along with many others) carry out a DFCA:

- For each time period, the company's **free cash flow** (roughly, profits)
- The **federal funds rate**, or an interest rate set by the **Federal Open Market Committee (FOMC)**, which is closely related to the degree to which future free cash flow is valued less than current free cash flow (i.e., the **discount rate**)

The value of any DCFA depends heavily on the ability to accurately forecast these and other quantities. Once I learned of the centrality of DFCAs to the success of GBB Investors, I knew I needed to dig into the way these quantities were typically estimated by your employees and identify whether this was being done effectively. In the following two sub-sections, I outline current practices and investigations I did into the efficacy of each.

Operating Cash Flow Projection

While other values are central to estimating future free cash flow for assets in general (e.g., **depreciation**, **amortization**), when projecting it for companies you ultimately need to estimate the degree of **demand** for the company's product: how much desire there is among consumers to purchase the good or service. The modal investor at GBB Investors relies on some quantitative modeling to project this (based on e.g., the number of competitors and historical sales), but often deviates from these estimates on the basis of what are widely called "informed gut feelings". That is to say, investors would deviate from the quantitative evidence based on their subjective evaluations of cultural and fashion trends, often informed by conversations with area experts. For instance, in one instance I observed an employee of GBB Investors get ahold of Jay-Z to ask him whether he thought a particular brand of athletic shoes would "catch on", and if so, whether it would happen before the end of the next fiscal quarter!

Over my six-month period of observing the employees of GBB Investors, I almost immediately began recording how often and when those valuing assets deviated from quantitative projections because of these "informed gut feelings" and then how these assets faired in the future. These "informed gut feelings" were used in a shockingly high 65% of valuations, with deviations from the forecasts ranging from small to large. However, in my retrospective analysis I found that a majority of these deviations lead to worse investment decisions than the quantitative projections would have. In sum, these incorrect intuitions I observed cost GBB Investors tens of thousands of dollars in just six months. If the employees had stuck to the quantitative modeling, GBB Investors would on average be more successful.

So why do the employees continue to rely on these "informed gut feelings" despite them leading the employees astray? Well, it seems to be the case that while losses experienced from these deviations are small and common, the gains from the occasional correct deviation are large and celebrated (but rare). This leads most employees to erroneously believe that "informed gut feelings" generally *help* the bottom line of GBB Investors, when in fact it only rarely does so and on average *hurts* the bottom line. When talking with an investor about these "informed gut feelings", they told me the following:

> *"We have to value companies that sell all kinds of products… Shoes, music, technology; you name it! There aren't a lot of common threads across all those different markets, which makes it difficult to find a rigorous and consistent way to make all those valuations. The one thing that is common, across all those decisions? Me! And my gut! I suppose if there were more indicators I could use across my entire portfolio to estimate demand, I'd consider using them."*

Federal Funds Rate Projection

Although there are many **macroeconomic forecasts** available to predict changes in the federal funds rate (which typically happens eight times a year) from various sources, employees at GBB Investors generally consider only one or two sources at a time. Over my six-month period of observing employees at GBB Investors, I've seen three different forecasts used by the employees. Specifically, for the first four months everyone used exactly two: one from the **International Monetary Fund (IMF)** and the other from an online financial blogger and bitcoin enthusiast, Money4Daze. Around the fourth month of my observation period, interest rates changed in a way inconsistent with what the IMF forecast had predicted. The employees remarked that the IMF projections had led them astray several changes to the federal funds rate in a row, and all agreed to stop using the IMF projections and to instead rely on a forecast generated by a Wall Street investment bank instead of the IMF projections. They continued (and continue still) to use the Money4Daze forecast as well.

When the two or more forecasts the investors are considering disagree, they typically perform a **sensitivity analysis**, in which they would test whether the two projections would lead to drastically different investment decisions. If they did not lead to different decisions, they chose the one that seemed more reasonable to them to make their exact valuation (assuming differences were negligible). If the two projections would result in drastically different investment decisions, then the employees typically would meet, discuss the merits of the different forecasts (and bring in their own subjective assessments of how the economy would fair), and then collectively decide on which forecast to follow.

I was interested in what would happen if, instead of relying on only one of the forecasts, the employees had relied on some aggregation of multiple forecasts. For my test, I simply took the average predicted federal funds rate from the three forecasts the employees had used: the IMF projection, the Money4Daze projection, and the investment bank projections. I found that this average substantially outperformed any of the individual projections. I then went back to my records and re-estimated investment decisions and profits had the employees used this average, and I estimate that GBB Investors would have performed much better.

The Human Element

I was surprised at how big a role human instincts and subjective evaluation were playing in decisions at such a large investment firm, so I talked with a couple managers about my findings. Managers were surprised to find that these instincts were leading investors astray: their impression was that this was the “secret sauce” of GBB. They pointed to specific examples of when investors made decisions which, on paper, seemed ill-advised, but had provided GBB with the gains it needed at pivotal times in the company’s strategic trajectory. One manager kept referring to the “human element”, or the allowance for GBB’s trusted and veteran employees to exercise judgement even in the face of overwhelming quantitative evidence. Indeed, as I noted above, there were times where these intuitions lead to surprisingly successful investment decisions. Overall, though, investors’ regular reliance on these intuitions seemed to be getting in the way of GBB’s success.

SECTION II: Proposed Solution

I’m proposing to incorporate contemporary analytics into estimating the operating cash flow of companies as well as the federal funds rate. In general, I believe that predictive algorithms with access to the right kind of data will be able to estimate these better than the employees of GGB are today with their current practices. However, I also recognize that the “secret sauce” of GBB Investors has been the “human element”, and I want to make sure GBB can continue to harness that to its advantage in the rare instances where it might pay off.

Therefore, I propose solutions that incorporate contemporary analytics but reserve room for human judgement. However, given the track record I've observed of employees' discretion, I think it's important to limit the degree to which employees can deviate from quantitative prediction. This will involve implementing an **algorithm/human decision-making model** that considers tradeoffs. In each of the two sub-sections below, I propose—for operating cash flow and the federal funds rate respectively—a way to use contemporary analytics to improve GBB's projections and then an algorithm/human decision-making model that will optimize the use of these predictions. My goal is to curb the negative influence of your employees' incorrect intuitions, while still harnessing the power of the "human element", which has provided a competitive advantage (at times) for GBB Investors.

Predicting Demand from Social Media

Social media is an increasingly important part of society. Twitter, one of the largest platforms that also allows the public to **scrape** its data (i.e., get access to data stored on their private servers) with relative ease, has a projected 330 million monthly active users.[1] Each of those users is a potential consumer for many of the products and services sold by companies that GBB Investors need to valuate. I propose, therefore, to use **automated text analysis** (the use of computers to extract quantitative information from written language) applied to social media to predict future demand. Specifically, I propose that we use the Twitter **API** (application programming interface: the protocol and software by which we can extract information from a website) to collect tweets that mention companies GBB Investors are valuing, use automated text analysis to measure "public sentiment" towards each company and its products, and use this to predict future sales.

For the automated text analysis portion of this prediction task, I'm proposing the use of **term frequency analysis**, which we can use to quantify how generally "positive" or "negative" the overall valence of a piece of text is. These techniques generally count the frequencies of empirically derived collections of words in documents, which can then tell us something about the tone of the text. There are many possible specific analyses one could implement, but this issue is beyond the scope of this report (though in brief I would suggest measuring the positive and negative emotion components of **LIWC**). When evaluating a company, the positivity and negativity of tweets about the company and its products can be used to predict free cash flow the following quarter.

Here's how we could test the viability of this proposal. First, we select a large set of companies representative of those GBB Investors might be interested in investing in—for instance, Nike Inc (NYSE: NKE). Then, for each company, we query the Twitter API for all tweets mentioning "Nike" or one of its products. Then, for each tweet, we quantify its tone via term frequency analysis. Then we take the average tone of tweets posted each financial quarter. Now, we have a single measure of "public sentiment" towards Nike for every quarter. I'm proposing that public sentiment towards Nike this quarter will predict sales for Nike next quarter. So, if public sentiment

[1] https://financesonline.com/number-of-twitter-users/#:~:text=Demographics%2C%20Breakdowns%20%26%20Predictions-,Number%20of%20Twitter%20Users%202022%2F2023%3A%20Demographics%2C%20Breakdowns%20%26,the%201st%20quarter%20of%202019.

towards Nike decreases today, we should expect a drop in sales—and a subsequent drop in free cash flow—next quarter. If we observe in historical data that public sentiment reliably predicts future sales as a I expect, then employees of GBB Investors can and should use this data to better predict demand, free cash flow, and ultimately free market price.

For this decision-making process, I'm recommending the use of what is known as an "**algorithm-as-advisor**" model. Specifically, when an investor is estimating free cash flow for a company, the estimate of future demand from the automated text analysis should be given to them as one piece of information to consider. Ultimately, however, the final decision should be left up to the investor. This is important because future demand can be influenced by many different factors, and public sentiment towards the company (and even future demand—if we could directly assess that) is only part of the equation. Further, this allows us to maximally harness those "gut feelings" while also providing a consistent and data-driven input that investors can learn from. Other algorithm/human decision-making models are especially good at curbing various forms of human **bias**, but on the whole I don't see a good reason to suspect that is what is leading to suboptimal decision-making. Therefore, if the public sentiment analysis does provide valuable information, I expect the investors to learn as much and rely on the results of the analysis more as time goes on.

Predicting Macroeconomic Conditions from an Ensemble of Projections

As mentioned above, there are many available macroeconomic forecasts. But which one should be trusted? As I suggested, by combining information from multiple forecasts, we can predict macroeconomic conditions (and therefore the federal fund rates) more accurately than any of the individual forecasts by themselves. I specifically suggest that we combine these individual forecasts via **supervised machine learning**, in which we train an algorithm to predict one variable (an **outcome**) from other variables (**features**) based on data. The basic intuition is that we will train artificial intelligence to combine these different forecasts into a single "super forecast". While there are many possible algorithms one might use and deciding on a particular one is beyond the scope of this report, my initial suggestion would be to use a **random forest** model.

The first step to this process would be to collect a **dataset**, or a collection of records about outcomes and features. In this dataset, the **observations** would correspond to each time the FOMC changed the federal funds rate. Each observation would be associated with an outcome—what the federal funds rate was actually set to—and features—what various forecasts predicted the federal funds rate would be at some set time before the rate was set. So, for instance, one observation would correspond to May 2004. The outcome for this observation would be equal to 0.99% (the actual federal funds rate at that time). The features would be, for each forecast available to GBB Investors, the prediction that forecast made in December 2003 (sixth months before May 2004) of what that the federal funds rate would be in May 2004. Separate datasets could be collected to train models meant to predict the federal funds rate at different time horizons (e.g., one dataset containing predictions six months ahead of time, one containing predictions one year ahead of time, etc.).

The second step is to divide our dataset into a **training set** and a **test set**. That is, we will use most of the data as examples our algorithm will learn patterns from, but other data should be reserved and instead used to evaluate how well the algorithm has learned those patterns in the data. If we used data that the algorithm learned from to also evaluate it, we might be misled about how well the algorithm learned to recognize patterns in the data. Specifically, the algorithm could "**overfit**" to the data, or learn its idiosyncratic characteristics that will not generalize to new data, and we would mistake this overfitting for real and effective learning. Evaluating the algorithm on a test set avoids this issue.

The next step is to **train** and evaluate the algorithm. The specifics of how this is done are beyond the scope of this report, but this should be carried out by a machine learning practitioner (ideally somebody with a computer science degree from some fancy school like Stanford). Importantly, the individual(s) who train the model should also report how well the algorithm predicted the data in the test set (the data that was reserved for evaluation). This will give you a sense of how well the algorithm will predict new data in the future. While discussing the different possible metrics they might report is beyond the scope of this report, I would recommend learning what the **mean absolute error (or MAE)** is. This tells you, on average, how wrong the algorithm's predictions were in the test set. For instance, if the real values in the test set were 3.22%, 1.35%, and 1.13% and the algorithm's (respective) predictions were 3.42%, 1.15%, and 1.33%, the MAE would be 0.2% (since each guess was either 0.2 above or 0.2 below the actual value).

In this case I believe a so-called "**justified deviation**" model, in which the algorithm makes preliminary decisions that can only be overturned by an investor if they justify doing so to an anonymous group of reviewers, will be especially effective. Specifically, each investor will be given the federal funds rate as predicted by the algorithm and expected to use that in their valuation decisions. If they believe, for whatever reason, they should adjust this estimate, they must submit their reasoning to an anonymous panel of fellow investors. If the anonymous panel decides that the deviation is reasonably justified, the investor can use their desired estimate. Otherwise, the investor should use the prediction generated by the algorithm. I think allowing an investor to possibly deviate from the model's predictions is important, so as to retain the "human element" that has made GBB successful, though overriding these data-driven projections should only be done rarely, through reasonable effort and care, and when the investor has a clear and clever reason for wanting to do so.

SECTION III: Potential Drawbacks

While I believe my solution could help GBB be even more successful than it already is, I also recognize that my solution comes with risks and drawbacks. In the spirit of transparency, I'll detail those potential risks and drawbacks in the following section.

The Limits of Social Media, Sentiment Analysis, and Seeing the Future

Recall that I proposed to use Twitter data to predict demand for a company's goods or services via term frequency analysis.

One could reasonably be concerned that **what we observe on Twitter is not representative of the market as a whole**. That is, those who are on Twitter might not be representative of all possible consumers of a company's goods or services. Further, using data at the level of tweets might bias the results to reflect the attitudes of those most prolific on social media, who might be even less representative of the general public than all Twitter users. This is a fundamental issue of leveraging social media data in the financial technology sector.

A second reasonable concern is that, even if we could collect how every possible consumer evaluates a company, sentiment might not be entirely representative of demand. For instance, even though most people speak negatively about Amazon, they still use their online retailer store over competitors who are seen as more virtuous and talked about more positively on social media. **Consumer's demand comes from myriad sources**, so a simple measure of tone used to discuss a company or its products might not be entirely predictive of its future sales.

A third concern is that social media couldn't possibly provide us a strategic advantage over other investors since the data is publicly available. According to traditional economic theory, **the price mechanism** incorporates all available information, so how could we capture significant value from information that is so widely available? Only time will tell whether this concern holds water, but it is a reason to doubt the viability of my solution *a priori*.

Finally, there is a looming moral concern with my solution. Companies, especially large ones, are aware that how they are portrayed on social media can influence how the public thinks of them and their products. Because of that, we might expect these same large companies, who have room in their budgets for non-traditional advertising, to artificially increase how positively they are discussed on social media sites such as Twitter. For instance, companies regularly pay social media influencers to talk positively about their products. Whether or not this increases the public's awareness of the company's products (the intended effect), this would influence our analysis since the influencers themselves are on social media. That is, along with genuine feelings potential consumers have about products, we will inevitably end up including in our analysis opinions that were paid for. Insofar as this analysis of public sentiment guides investment decisions, we may end up rewarding companies simply because they enough money for large-scale, non-traditional advertising. This could lead to a viscous cycle in which we contribute to growing income inequality.

This could become even worse if companies *realize* investment decisions are being made based on tone of social media content. If that happened, companies could begin creating "bots" (users on Twitter who do not correspond to a real or natural person and are algorithmically controlled) to post positive content about their company with the intent of influencing our investment decisions, thus **gaming** our measure. At this point, the data from Twitter would become highly unreliable at predicting future sales (as per **Goodheart's Law**: "when a measure becomes a target, it ceases to be a good measure"). Perhaps more worryingly, we would end up rewarding companies for activity that in no meaningful way contributes to the economy. This would immediately hurt our clients and, inevitably, consumers writ large.

Getting Lost in the Random Forest

Recall that I proposed to use machine learning to combine the various forecasts investors used to estimate the federal funds rate into a "super forecast", which would outperform any of the individual forecasts. Then, I suggested that investors be bound to these predictions unless they could convince an anonymous panel of other investors that their justification for deviating was reasonable.

One immediate concern is that such a system constrains the "human element" too much. Even given the analyses I presented above, one might argue that part of the **organizational culture** of GBB was the leeway it gave to its investors. If I take that away with the solution I've presented, I might disrupt what has been to date a winning formula. This is a fair critique of my analysis: I analyzed each investment decision as if they're independent, when in fact the processes by which these decisions are arrived at might be part of a "package deal". If that's the case, interfering with one element (even though it seems dysfunctional in isolation) might have unexpected and disastrous consequences. There is of course no way that we can now whether this is true or not before implementing the solution, making it inherently risky.

Another reasonable concern is in the use of the justified deviation model (where investors must convince a panel of other investors anonymously if they want to deviate from the algorithm's recommendations). Specifically, even though the panel is anonymous in principle, investors as a group might coordinate so that they aren't ever forced to abide by the algorithm's decisions. In that case, my solution becomes far less useful, since the investor's preferences will override the algorithm's, even if the justification isn't reasonable. There would still be some benefit, though the decision-making model would essentially reduce to an algorithm-as-advisor model, which I described earlier in this report. One might monitor how often deviations are considered justified by the panel and perhaps even implement periodic reviews from managers, but there's no easy solution for overcoming this.

CONCLUSION

In conclusion, I've identified practices at GBB Investors that I believe contribute to sub-optimal investment decisions. I observed these practices during a sixth-month observation period, during which I also formulated a unique set of solutions that uses contemporary analytics to produce data-driven insights into the fair market value of assets. While my proposed solutions come with risks and drawbacks, I believe that they—on the whole—will greatly help GBB Investor's bottom line.

I'd like to thank the managers of GBB Investors for this opportunity and for being gracious hosts during my time there. I certainly learned a lot about your organization in the process, and I hope that the recommendations I make were worth my exorbitant consulting fee. I'd also like to thank the students of my People Analytics course, who have made this summer truly rewarding, challenging, and awesome.

Best,
Austin van Loon