

Case Studies:
Do You Know the Provenance of Your Alternative Data *

Seoyoung Kim
Associate Professor of Finance
Santa Clara University

June 2019

[For *Journal of Investment Management*]

“Case Studies” presents a case pertinent to contemporary issues and events in investment management. Insightful and provocative questions are posed at the end of each case to challenge the reader. Each case is an invitation to the critical thinking and pragmatic problem solving that are so fundamental to the practice of investment management.

* The author may be reached at: srkim@scu.edu; 500 El Camino Real; Santa Clara, CA 95053.

Do You Know the Provenance of Your Alternative Data?

In her continued assessments of fund P&Ls, the managing principal of a major quant fund has decided to redirect more resources to its alternative data efforts, which have shown much greater profit margins than traditional data use. So-called “alternative data”, which has evolved to entail the culling, cleaning, and curing of an unstructured mass of information to produce structured data, is being used by the fund to glean predictive insights that can generate more inventive and profitable investment strategies.

Tom, who oversees the fund’s alternative data strategies, is now tasked with seeking out ever newer alternative data sources and overseeing the creative direction and backtesting of the new data in various trades. One of Tom’s early successes in this sphere entails his decision to purchase data gleaned from parking-lot satellite imagery, which he used to develop a profitable trading strategy involving firms in the retail industry. Another entails his decision to purchase sentiment and deception scores which were produced from applying natural language processing (NLP) techniques on earnings call transcripts. But the business of alternative data has been expanding rapidly, and there are now even more interesting and exciting choices.

For instance, one data provider is offering structured geolocation information gleaned from cell phone apps. Tom envisions using this dataset to derive a more finely tuned predictive metric for his earlier successful trading strategy of firms in the retail industry. He also envisions using this information to develop trading strategies based on work-force movement and productivity issues in labor-heavy industries, such as in real-estate development. Another data provider is offering predictive sales metrics based on information gleaned from credit card transactions, which Tom views as a valuable complement to the geolocation data in improving his retail-firm trading strategy.

The managing principal is impressed by Tom's proposals, but remains guardedly optimistic. For one, she is concerned about the provenance of the geolocation data. That is, does Tom know from which apps the geolocation information was gathered? And was consumer consent legitimately obtained? She vividly recalls a case in which an app developer offered a "free" flashlight application without appropriately disclosing that users' location information would be sold to third parties. Ultimately, the Federal Trade Commission (FTC) charged the developer for deceiving consumers, and she wonders what might be the fund's exposure as a sophisticated end user of such information.

Furthermore, continued profit margins depend on exclusivity of the data, which contributes to a higher risk of violating insider trading and privacy laws. That is, if non-public information, such as that gleaned from credit card transactions and geolocation data, can produce high-alpha investment strategies, should the fund be concerned that it is purchasing *material non-public information* (MNPI)? Thus, seemingly profitable strategies may not be so profitable once these risks are appropriately factored.

With that, Tom and the managing principal pack up for the day with the view to meet with a special regulatory council at the SEC along with their own legal counsel before deciding how to proceed.

Questions

- How much responsibility should the fund assume in investigating the provenance of the alternative data it is thinking of purchasing?
- What are key elements the fund should consider in assessing whether the alternative data might be considered *material non-public information*?

- What other non-legal ethical concerns may arise from the use of alternative data in investment management?