

The Power of a Nobel Prize

Xi Chen

October 16, 2017

Course: MACS 30000 Perspectives on Computational Analysis

Assignment: Proposing an Observational Study

Instructor: Dr. Benjamin Soltoff

Introduction and Research Question

On October 9, the 2017 Nobel Prize in Economics was awarded to Richard H. Thaler, one of the founding fathers of behavioral economics and a professor of the Booth School of Business at the University of Chicago. The other two founding fathers of behavioral economics are Daniel Kahneman and Amos Tversky, who published the paper “Prospect Theory: An Analysis of Decision under Risk” in 1979. Daniel Kahneman was awarded the Nobel Prize in Economics in 2002 for his contribution to behavioral economics.

Behavioral economics has never been in the mainstream of economics before, because one of the leading goals of behavioral economics is to challenge the cornerstone of traditional economics theory: the assumption of rationality. However, after October 9, enormous attention to behavioral economics has been aroused among the public. The mass media around the world became eager to publish newsletters, interviews, or comment articles about behavioral economics. People are greatly interested in knowing more about behavioral economics. Behavioral economics has never been more popular than it is today, being intensely discussed by people from both industry and academia.

No one would deny that this is the power of a Nobel Prize. However, does the Nobel Prize just bring fleeting trend to the certain field? Does the Nobel Prize boost up long-lasting development in a certain academic field? Does the Nobel Prize increase awareness of a certain academic field among the public in the long-run? How large the impact of a Nobel Prize would have? To sum up, the research question in this proposal is what are the longitude and magnitude of the impact a Nobel Prize had on the public, industry, and academia in the long run.

Data Sources

For the impact on the public, I plan to collect data from Google Search. Past research indicates (Ginsberg etc, 2009), search engine query data can be useful to reveal public’s focus and interest. Data, such as how many times that people enter “behavioral economics” or related words into Google Search, is planned to be collected.

For the impact on the industry, I plan to choose two famous and popular news media, the Wall Street Journal and the Financial Times, which play important roles in the industry. Data, such as how many articles in these news media talked about behavioral economics, is planned to be collected.

For the impact on academia, I plan to collect data from the Google Scholar, in which I can track the number of papers that were published over a certain time period and in a certain field. In addition, I will record citations of the relative papers, which can be used to measure the magnitude of the impact.

Research Design

This research will be a longitudinal study. There are three important time points in the history of behavioral economics: 1979, when Daniel Kahneman and Amos Tversky published a paper which was considered as the foundation of behavioral economics; 2002, when Daniel Kahneman got the Nobel Prize in Economics; 2017, when Richard Thaler was awarded the Nobel Prize in Economics. These time points separate the complete data set into several subgroups: before 1979, between 1979 and 2002, between 2002 and 2017, and after 2017.

The research designs in this proposal are based on counting things and measuring things (Salganik 2017). Forecasting may also be included in the future research. By crawling and scraping data from the sources mentioned above, I will get a large scale of data, and then a lot of descriptive statistics can be done to examine how people react to the Nobel Prize. For example, by visualizing the quantity of the relative papers in different time groups, I can see the developmental trend and the continuous changes along the time. Thereafter, I can examine if there is a casual relationship between the Nobel Prize and the academic development. I can also analyze data to examine whether the differences in various time groups are significant. For the magnitude of the impact, with the number of citations that the relative papers have, I can test the question like is a paper with more citations more likely to be published before or after the Nobel Prize was awarded? After the descriptive statistics, I will progress to inferential statistics, and try to build a model which can predict the future development of behavioral economics after 2017 Nobel Prize was announced.

Good Characteristics of Big Data

Big

The data sources mentioned above include an enormous scale of data. For example, people use Google Search every day, so the search engine query dataset can be so large. In addition, this is a longitudinal study. Between 1979 to 2017, numerous newspapers and academic papers had been published, so there are a few potential relative articles that can be investigated.

Always-on

As Salganik (2017) mentioned in “Bit By Bit: Social Research in the Digital Age”: “Always-on big data enables the study of unexpected events and real-time measurement.” The Google search engine collects data all the time, so it records the instant changes before and after the Nobel Prize being awarded. Similarly, the news media publish articles every day, and never stop unless under some special circumstances.

Non-reactive

Participants don’t know their search engine query data would be collected when they enter words into the Google Search, so they don’t react to the researcher’s measurement.

Bad Characteristics of Big Data

Inaccessible and Sensitive

The data from the Wall Street Journal and the Financial Times can be purchased by signing up for their membership, so these data are not hard to get access to. Google scholar data can also be attained by verified academic affiliated accounts. However, the Google search engine query data may be inaccessible because these data, considered as business secret, would be controlled by the company. Personal search data are so sensitive that the company worries a lot about the risk of data breach. Hence, as Salganik (2017) suggests: “Google—and most large companies—are very risk-averse about sharing data with researchers.” If I cannot

get the data from the Google company, I will consider collecting data from the social media and social networking platforms like Facebook and Twitter, which can also reflect the public's focus.

Dirty

The big data are always dirty. When crawling data by a designed algorithm, some of the papers may be flagged as “highly relative”, but in fact they can be irrelevant to the topic that the researchers want to investigate. For example, many people have experienced such a situation: when you enter several key words in the Google Scholar, some the papers showed up may not be the ones that you tend to find. Therefore, it is hard to get the completely clean data. To deal with this issue, I plan to use content analysis techniques or advanced crawling algorithms to increase the efficiency of collecting clean data.

Reference

Salganik, Matthew J. (2017). *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press. Open review edition.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society*, 263-291.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012-1014.