# PS1

**Yuqian Gong**

## Part 1

**Problem 1. Describe how to access data, where it is stored, who curates it. Make sure to use the original source and curator in addition to the NBER site to which I have linked.**
The dataset come from The NBER U.S. Patent Citations Data File. It is stored on NBER website and can be downloaded freely. It has data sources, the first is Patent Oce and the second is Compustat. The curators are the authors of the paper *The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools* who describe the data in detail including Bronwyn H. Hall, Adam B. Jaffe, Manuel Trajtenberg.

**Problem 2. Cite other key papers that have used this data.**
The data has been used to study the impact of technological diversity and alliance organizations on firm innovative performance (Sampson 2017). The data has also been used by researchers to show that positive technology spillovers on social scale dominate negative business stealing effects from product market rivals on individual firm scale(Bloom, Schankerman, Reenan 2013). Researchers also study how technology transfer within U.S. multinational firms changes using this dataset.

**Problem 3. Describe how the data were collected.**

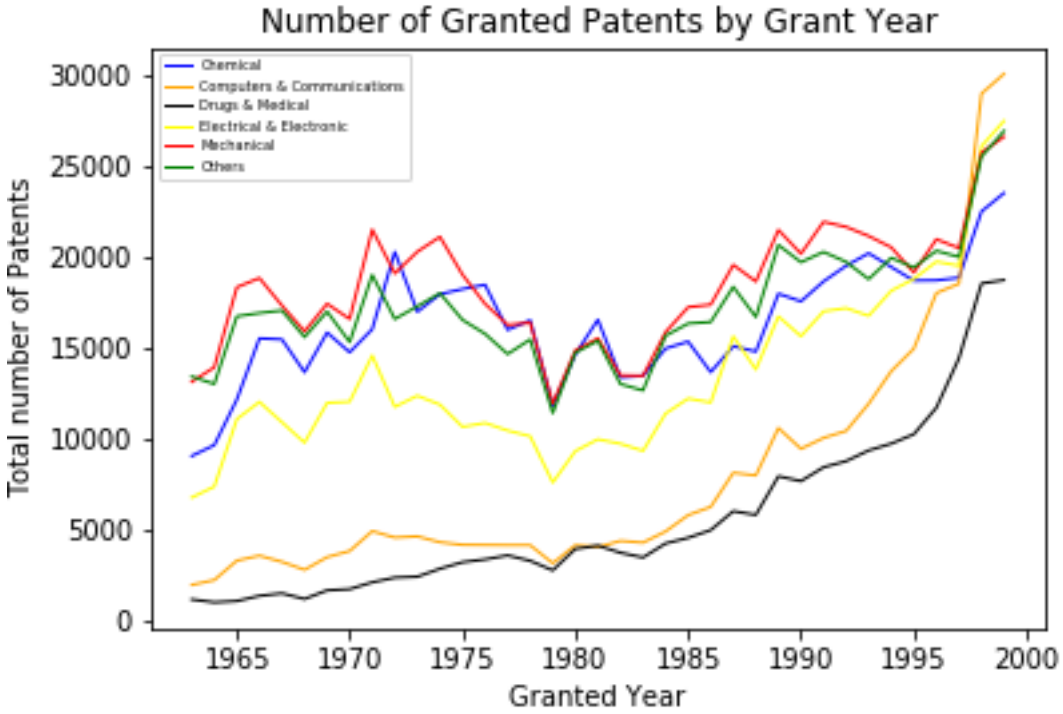
The data mainly come from the Patent Office, MicroPatent and constructed by the curators. The main dataset consist of all the utility patents granted during the period from January 1, 1963 through December 30, 1999, including the 10 orginal variables from the Patent Office and 10 constructed variables created by the curators. Another dataset include all 16,522,438 citations made by patents granted in 1975 to 1999, which come from the source of Patent Office and MicroPatent.

**Problem 4. Include a table that gives descriptive statistics for at least 8 key variables (you can do more).**

| | count | mean | std | min | max |
|---|---|---|---|---|---|
| Grant Year | 2923922.0 | 1983.55 | 10.98 | 1963.0 | 1999.00 |
| App Year | 2699606.0 | 1983.11 | 10.13 | 1901.0 | 1999.00 |
| Num of Claims | 1984055.0 | 12.08 | 10.27 | 1.0 | 868.00 |
| Num of Cmade | 2139314.0 | 7.72 | 9.00 | 0.0 | 770.00 |
| Num of Creceived | 2923922.0 | 4.78 | 7.35 | 0.0 | 779.00 |
| Measure of Generality | 2240348.0 | 0.32 | 0.28 | 0.0 | 0.94 |
| Measure of Originality | 2042151.0 | 0.35 | 0.28 | 0.0 | 0.95 |
| Mean forward Citation Lag | 2074641.0 | 8.31 | 5.80 | 0.0 | 96.00 |
| Mean backward Citation Lag | 2088785.0 | 14.10 | 11.77 | 0.0 | 154.00 |

From the table, we can see that the number of observations is 3 millions. The available patent granted year range is from 1963 to 1999 and the available patent application year range is from 1901 to 1999. The number of claims are the blocks of the patented invention and they are indicative of the 'width' of the invention. The average number of claims is 12 and the standard deviation is large, which is about 10. The number of citations made may indicate how related a patent is to others. The number of citations received may indicate how important and popular a patent is. The average numebr of citations made are greater than the average number of citations received. The generality score measures how widespread a patent's impact is and the originality score measures how widespread cited patents' impacts are for a patent. The average scores for them are pretty low actually. Mean forward citation lag is the average difference in years between one patent and patents citing it. Mean backward citation lag is defined as the average difference in years between one patent and patents it cited. Noticably, the mean forward citation lag is much lower than the mean backward citation lag, with 8.31 years compared with 14 years. This is generally because that most patents are not cited and the average year differences on patents with many uncited ones will yield a lower difference.

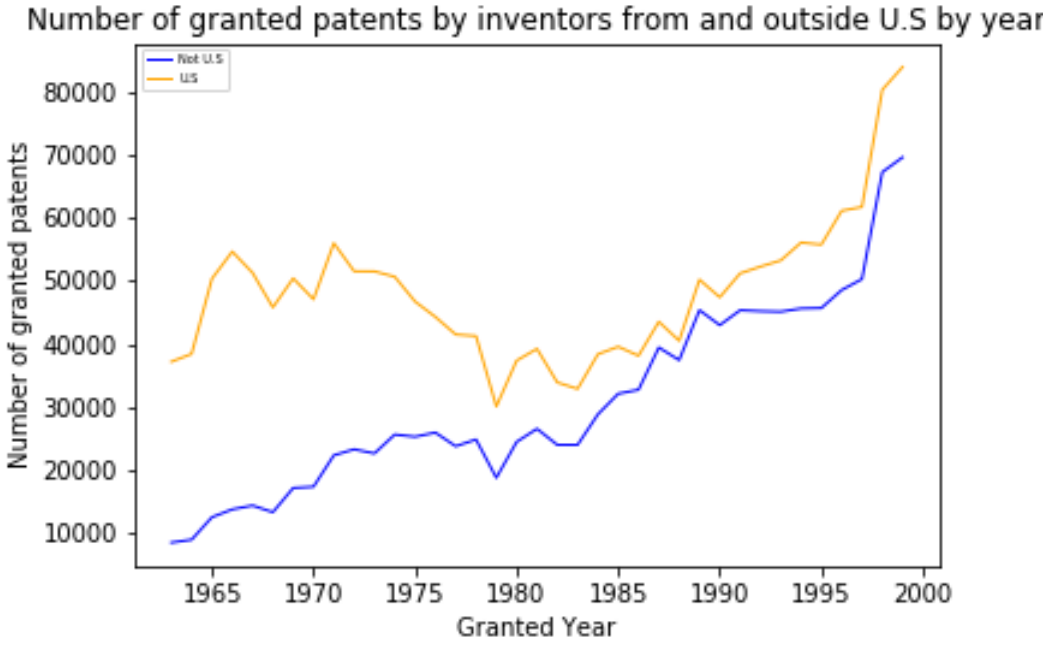
**Problem 5. Include at least one key visualization of the data that exhibits an interesting characteristic.**




I drew a graph showing the trend of number of patents granted in each technological category by year. It is interesting to see that in general over time, mechanical industry has the most patents granted each year while drug and medical industry has the least. Around

1980, the number of patents granted in each category had a downward trend and the number headed up after mid 80s. Besides, it is not hard to find that the increase in the number of patents granted in Computers and Communications industry are most remarkable over the 40 years. From about 2000 each year in 1963 to about 30000 each year.

**Problem 6. Show at least one conditional (slice) description of the data (e.g., all variable descriptive statistics by nationality of survey respondent). This can be a table or visualization.**




The graph above shows the trend of granted patents from inventors inside and outside U.S. over time. It is clear to see that the number of granted patents from inventors outside U.S has gradually caught up during the thirty-five years and they share similar trends after 1975.

# Part 2:

### 1. State the research question of your assigned paper.

The paper tests two types of predictions that is from a recent theoretical work on the provision of advice potential customers. Firstly, the first set of predictions is about studying the impacts of commissions on the quality of advice provided by agents. The second set of predictions study how regulation and customer types affect the quality of advice.

### 2. What data did the paper use?

The data are collected from trained auditors involved in three experiments in Indian cities. The first experiment tests whether agent advice is based on customers' needs or customers'

beliefs about products (evaluating the quality of agents' advice). The second experiment studies whether agents respond differently to different types of customers( evaluating the impact of customers' sophistication). The third experiment studies how disclosure regulation affects the quality of advice that insurance agents provide (evaluating the impact of disclosure). The data size are 557, 257, 217 respectively.

**3. What theory did the paper reference in order to interpret the data? (Note: it is possible that the paper has no reference to theory.)**

The theories are based on the provision of advice to potential customers. Part of the models predict that some consumers will get low-quality advice to purchase a complicated product that produces higher commissions for agents( Inderst & Ottaviani, 2012c; Gabaix & Laibson, 2006). Also, to explain the impact of customers' sophistication on advice quality, the models by Inderst and Ottaviani(2012) showed that sophisticated customers will receive better advice. They also showed that disclosure requirement on commissions can improve advice quality by converting unaware customers to customers with knowledge on commissions.

**4. Was your assigned paper a descriptive study, an identification exercise, a numerical solution to system of equations study, or some combination of the three? (These are the three classifications we discussed in class.)**

I think the paper are both a descriptive study and identification exercise. Some descriptive facts on the data collected from trained auditors such as the statistics summary of table 2 in the paper. Beyond that, the main part of the research is identifying what the quality of insurance agents' advice is and how disclosure on commissions and customer sophistication impact the quality of advice.

**5.What computational methods did this paper use to answer the research question? What was their result or answer to the question?**

The computational methods include OLS and GLS, which included dummy variables and incorporated location fixed effects.

**6. Think of yourself as an academic referee. Give two suggestions to the author(s) of your assigned paper of things the authors might do to improve their results or strengthen their evidence for the answer to the question.**

I think the randomized control treatment can be conducted in stratified groups such as study the treatment effects on agents with different education background, different income levels and etc. Besides, the large percentage of agents owning whole insurance cannot imply that they actually personally believe that whole insurance are better products. Their general income might affect their personal choice or they get great discounts to buy whole insurances as benefits from their company; therefore it is worth further studying the reasons why the personally prefer whole insurances.

Citations:

Rachelle C. Sampson. “R&D Alliances and Firm Performance: The Impact of Technological Diversity and Alliance Organization on Innovation”. *Academy of Management*, November 2017, Vol 50, No.2, 364-386.

Nicholas Bloom, Mark Schankerman, John Van Reenen. “Identifying Technology Spillovers and Product Market Rivalry”. *Econometrica*, July 2013, Vol 81, No.4, 1347-1393.

Lee G. Branstetter, Raymond Fisman, C. Fritz Foley. “Do Stronger Intellectual Property Rights Increase International Technology Transfer? Empirical Evidence from U.S. firm-level Panel Data”. *The Quarterly Journal of Economics*, February 2006, Vol 121, No.1, 321-349.

Inderst, Roman, and Marco Ottaviani, “Competition through Commissions and Kickbacks,” *American Economic Review* 102 (2012a), 780–809. ——— “Financial Advice,” *Journal of Economic Literature* 50 (2012b),494–512.

Del Guercio, Diane, Jonathan Reuter, and Paula A Tkac, “Broker Incen- tives and Mutual fund Market Segmentation,” NBER working paper 16312 (2010).