ORGB 671: Assignment #4
US Patent Office: A Causal Model

Group 4:

Will Stephenson	261082056
Hugo Garcia	260791363
Liliana Tretyakova	261086215
Emery Dittmer	260658030

1.1. US Patent Office

The United States Patent Office (USPTO) USPTO advises the president of the United States, the secretary of commerce, and U.S. government agencies on intellectual property (IP) policy, protection, and enforcement; and promotes the stronger and more effective IP protection around the world according to their website.

The USPTO is a large employer of patent examiners. In this basic study we are examining the human metrics of these patent employees. This includes turnover rate, mobility within the company and more. To accomplish this we are using a dataset provided by the USPTO (simplified by our instructor). Additional details can be found here and the datasets here.

For this assignment our group researched how the examiner composition of art_unit by gender, and art_unit at time t, affects likelihood of examiner transitions to another art_unit at time t+1. Using quarterly data to get the number of examiners per technology center, we set out to demonstrate causal validity.

First, we examined the number of examiners per quarter and year looking at the Technology Center data based on who will leave at time t+1. Then we plotted line graphs for each Technology Center examiner and observed similar trends across all Technology Centers for the number of examiners per Technology Center over time broken down by gender.

Next, we chose to compare two Technology Centers - #1600 & #2100 - and see the number of examiners at time t and time t+1. We chose these two centers visually based on the observed parallel trends from our plotted line graphs.
We determined the treatment period (t --> t+1) visually by seeing when the trends stopped being parallel, and the mean values became statistically different.

With our treatment period determined, we ran a causal analysis using the CausalImpact package in R and computed a 99.9% probability of a causal effect on our target variable, with a relative effect of 743%.

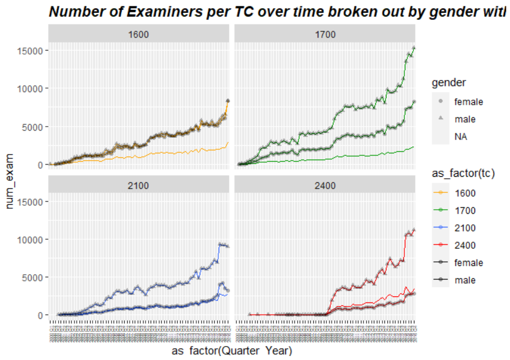
3.1. General Assumptions

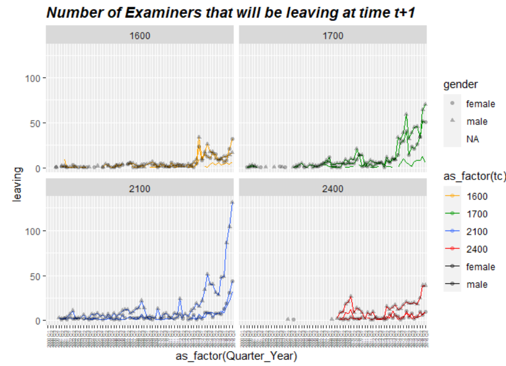
While investigating the USPTO data we are evaluating this data set with the assumption that for the observed data that there is some treatment A and a set of treatment covariant. We will Investigate both the size in number of units and the number of people leaving at time t+1 as the observations. The quarter – year and technology center (similar to a company division) are therefore covariates. With limited data and context we will evaluate the potential effects with some degree of uncertainty.

3.2. Stable Unit Treatment Value Assumption (SUTVA)

The SUTVA condition assumes that there is no interference between units. The treatment effect of each technology center should not effect the other. While we did assume that the two Technology centers are comparable, this is an assumption that backs all of our causal inference. While we do see parallel trends between these two centers there may be external factors that aren't visualized in the data that could affect what we are studying. The example in class of a tornado possibly hitting only West Virgina and not being factored into the examination of fiscal policy could be happening here. However, based on the available information that we have accessible to us, and the scope of this paper, we believe that these two centers stand in enough similarity to be used for a causal inference.

Further we are assuming independence in the two centers for treatments. Each technology center will not have a connection to the others as although they are in the same organization there is some separation with the leadership, hiring requirements and specialized knowledge. Their treatment should be independent from each other as their modes of operation are not interdependent. Based on the graphs for the number of employees and number of people leaving at time t+1 there are similar trends for all TCs except for 2100. Since we will only compare 1600 and 2100, the treatment of 1600 will stand as the control while 2100 will stand as treatment A. Therefore, there is only one version of a treatment as per the causal assumptions.





3.3. Consistency

We will assume that the technology center 2100 will undergo some treatment A, that will produce a result observed in the dataset. Based on the graphs and comparison of the different averages we believe that there is sufficient evidence to demonstrate that 2100 from the year 2012 has a significantly different t+1 leave number than the rest of the group. While we do not know what treatment any technology center has undergone, we can demonstrate the difference between them. Therefore, we believe that the consistency assumption, or that some treatment A has caused effect Y holds for this dataset.

3.4. Ignorability

Ignorabilioty holds that given any pre-treatment conditions the treatment assignment is independent from potential outcomes. The examiners have specialized knowledge and work within technology centers because of their specialized work. This does indicate that the assignment to the treatment is not random. However, both tc 2100 and 1600 have a similar population distribution based on gender and ethnicity. Given similar distributions we can assume to some degree that the treatment is independent from any known pre-treatment conditions, except for their assigned technology center. Furthermore, the potential outcomes of any treatment as measured at time t+1 are not known at time t. Humans do not have the ability to predict the future with any degree of certainty. Therefore, this assumption largely holds. While the examiners are not randomly assigned to technology centers or treatments, we can safely assume that they are not assigned to treatments based on any knowledge of outcomes of pre-treatment conditions.

```
 Descriptive statistics by group
group: 1600
              vars   n    mean      sd median trimmed     mad  min  max range
Quarter_Year*    1 133   34.25   19.27     34   34.25   25.20    1   67    66
tc               2 133 1600.00    0.00   1600 1600.00    0.00 1600 1600     0
gender*          3 133    1.50    0.50      2    1.50    0.00    1    2     1
num_exam         4 133 2881.35 1881.86   2516 2810.72 2170.53    4 8333  8329
leaving          5 110    5.56    6.60      3    4.11    2.97    1   34    33
               skew kurtosis     se
Quarter_Year*  0.00    -1.23   1.67
tc              NaN      NaN   0.00
gender*       -0.01    -2.01   0.04
num_exam       0.37    -0.52 163.18
leaving        2.17     4.85   0.63
------------------------------------------------------------
group: 2100
              vars   n    mean      sd median trimmed     mad  min  max range
Quarter_Year*    1 129   33.25   18.69     33   33.25   23.72    1   65    64
tc               2 129 2100.00    0.00   2100 2100.00    0.00 2100 2100     0
gender*          3 129    1.50    0.50      2    1.50    0.00    1    2     1
num_exam         4 129 2261.26 2159.71   1357 1936.73 1727.23    2 9340  9338
leaving          5 117   11.42   19.17      5    7.00    4.45    1  131   130
               skew kurtosis     se
Quarter_Year*  0.00    -1.23   1.65
tc              NaN      NaN   0.00
gender*       -0.02    -2.02   0.04
num_exam       1.33     1.52 190.15
leaving        3.83    17.15   1.77
```

3.5. Positivity

The positivity assumption states that for every observation or examiner the treatment assignment was not determined. The treatment A of technology center 2100 is deterministic to some degree as each of these require specific domain knowledge. However, based on the assumption holds true from two perspectives. First, we do not have access to any data that would determine the technology center of an individual such as their background or experience. The tc is not deterministic and can be hidden from the data. Secondly, the examiners have mobility between art units. Some examiners change from art unit 1600 to 2100 and vice verca. Therefore, there is no specific deterministic criteria for who receives treatment A vs the control group.

4.1. Treatment Period

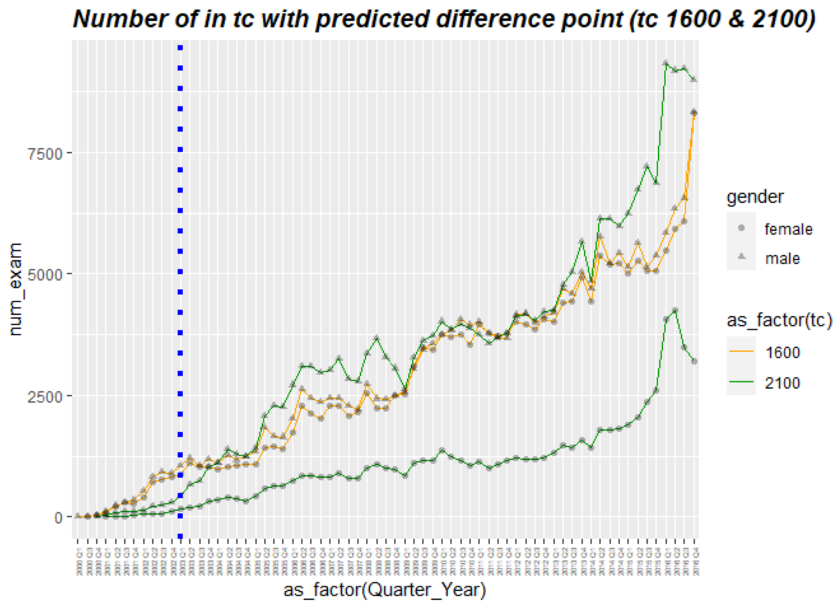


Figure 1: Number of Examiners in Technology Centers 1600 & 2100 with predicted difference point

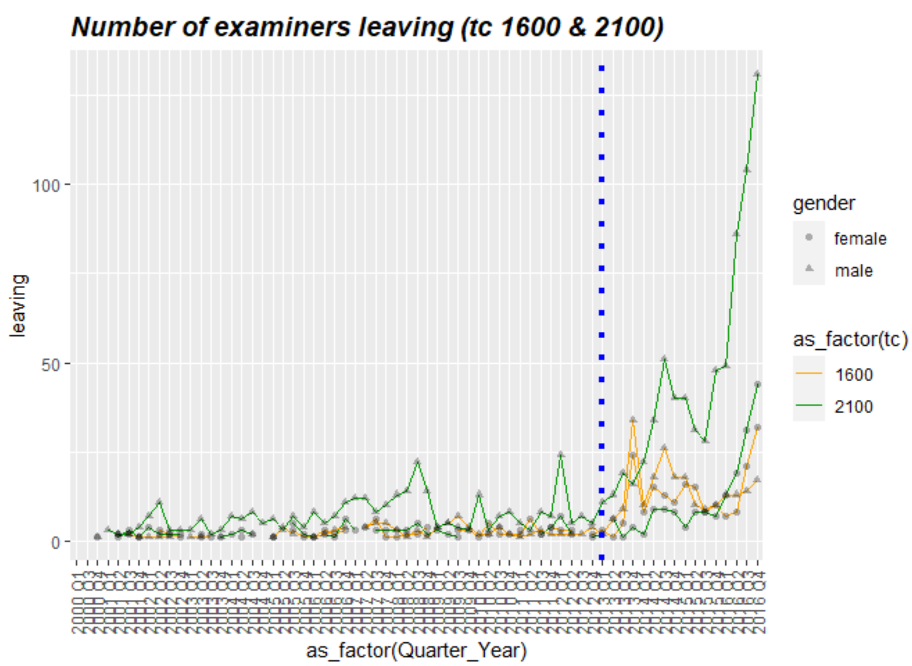


Figure 2: Number of Examiners leaving Technology Centers 1600 & 2100 at time t+1

4.2. Treatment Period

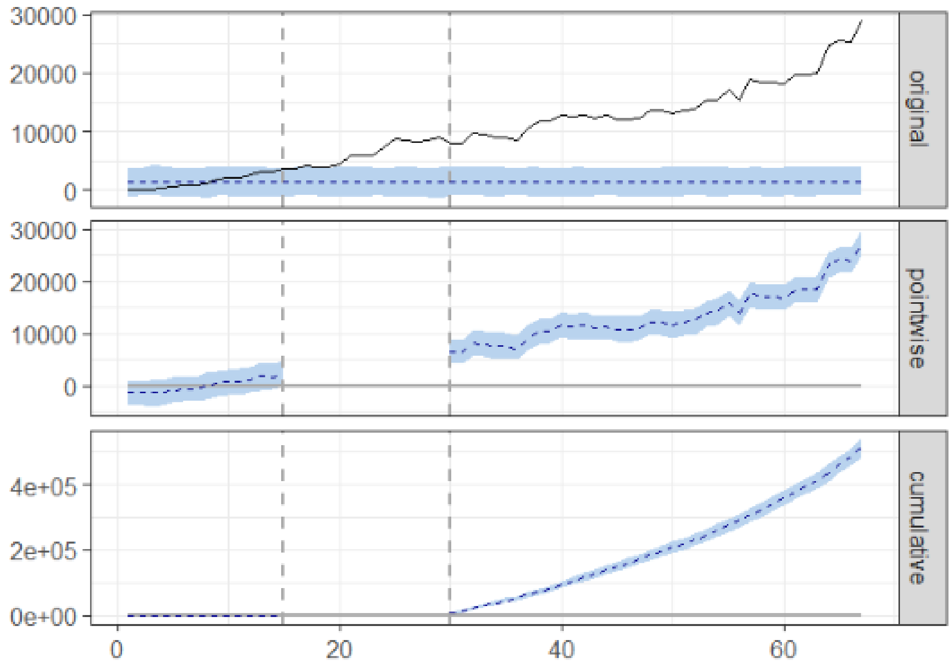


Figure 3: The Number of examiners in art unit 2100 considering a treatment period between Q3 2001 and Q4 2006 approximately.

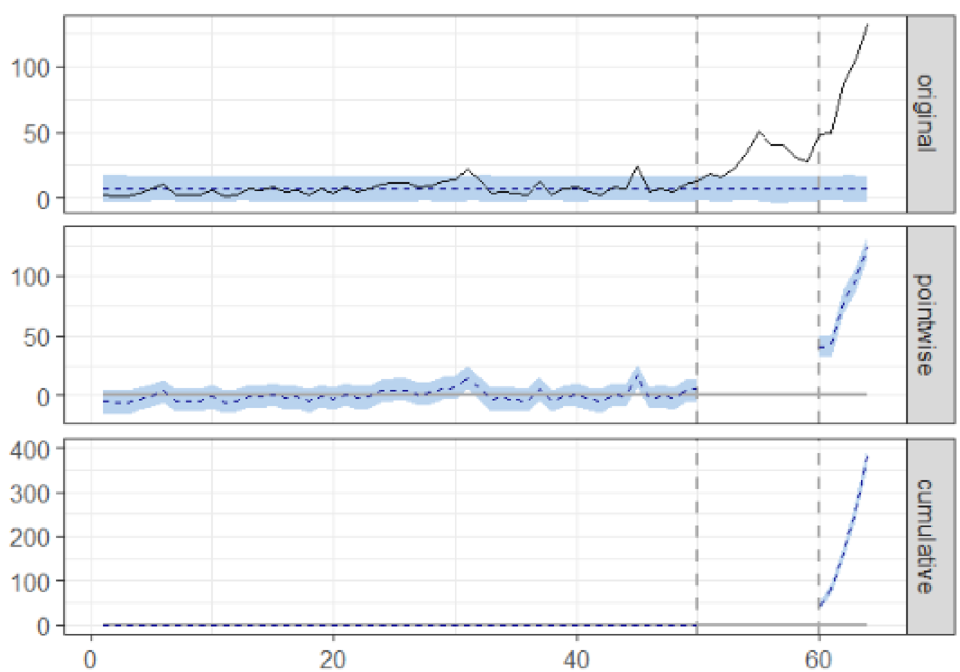


Figure 4: The Number of examiners leaving Technology Center 2100 at time t+1. Considering a treatment time of 2012 Q2 to Q3 2015.