Methodology for Analyzing

Matthew Borelli

ECP 4618: Research Methods for Studying Housing, Land and Cities

Dr. Samuel Staley

September 23, 2018

**Hypothesis**

The next step is to find a way to analyze local representation in films as it relates to state film tax incentives. Firstly, the measure of the local representation in film that I will use is a binary variable, called “local representation”, that measures whether a film’s primary state of production matches with the film setting. Primary state of production refers to U.S. state a film was primarily produced in and the film setting refers to the U.S. state in which the film’s story primarily takes place. Local representation will be classified as a 1 if the primary state of production and the setting state are the same, whereas a film’s local representation will be classified as a 0 if the primary state of production and the setting state differ. For the measure of state film tax incentives, the rating of states by tax incentive programs from Film Production Capital (2016) serves as a good measure of the “best tax incentive jurisdictions for production.” My hypothesis is that the rating of a state’s film tax incentive program has a positive effect on the amount of local representation in film for that state.

**Methodology**

Since the local representation variable is binary, it must be analyzed with a logistic regression instead of ordinary least squares regression. Logistic regression is a method that works under the same principles as ordinary least squares regression with the main difference that instead of predicting the actual value of the dependent variable, it predicts the probability of a binary variable equaling 1. The equation for the logistic regression is as follows:

1. logit (p) = β0 + β1X1 + β2X2 + β3X3 + β4X4 + ℇ
2. logit (p) = ln(p /[1-p])

Where *p* is the probability that the y variable (location representation) is equal to 1, *1 – p* is the probability that the y variable is equal to zero, and *ln* is the natural logarithmic function. Equation 2 is the explanation of what we are predicting: *p / 1 – p* is the odds of a 1 for the y variable, and applying the logarithmic function transforms it into a variable that spans all real numbers. Equation 1 is the regression equation that will serve as my predictor of the *logit* function.

The data for this analysis will consist of movies released from 2014 to 2016 that played in at least 1000 theaters in the U.S. The range of three years will ensure a good amount of data points, while the floor of 1000 theaters will ensure that the films played reached at least a fairly broad audience. There will be some restrictions on top of this to ensure that all data is relevant to state film tax incentives. First, only films that are set primarily in the U.S. will count for this to protect against skewing heavily towards films set in foreign countries or not the real world. On the same line of thinking, films primarily shot outside of the U.S. will not count either.

**Variables**

The previous sections have explained the thought process behind the left hand side of the equation, the logit function. However, attention must be paid to the left side of the equation to make sure the predictors should be included in the equation. The model for this analysis has five right hand side variables: X1, the independent variable; and X2-X5, the control variables. The table below provides a brief introductory description to each of the variables.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Name | Description | Expected  Effect | Source |
| X1 | Rating of state film tax incentive programs with respect to the production company | + | Film Production Capital |
| X2 | Percent of workers employed in the film industry in a state | + | Bureau of Labor Statistics |
| X3 | Population of the production state | + | United States Census Bureau |
| X4 | Binary variable that is set to 1 if the film is independent | + | IMDB |
| X~~5~~ | Listed production budget of the film | ± | Box Office Mojo |

To start, the independent variable for the analysis is the rating of state film tax incentive programs created by Film Production Capital. As aforementioned, Film Production Capital created a ranking of each state’s film tax incentive program that scales from zero to five based on how much of a tax break the film production company can receive. The higher the rating, the higher percent of tax breaks a production company can potentially receive on a project. The distribution of this variable will likely be left skewed once data is collected since production companies are more likely to film in states with higher ratings for their film tax incentives. This variable is expected to have a positive correlation with the dependent variable. Since films with higher ratings create larger film industries in the stay they are in, it is more likely that filmmakers form those states will want to create movies with their home state in mind.

The first control variable measures the percent of workers in a state that are in the film industry. This variable will be calculated using data from the Bureau of Labor Statistics by taking the number of film workers in the state and dividing it by the number of total workers in that state. The percent of workers in a state that are in the film industry should have a positive impact on the amount of films that display location representation for a similar reason as the independent variable. States that have a high percentage of workers in the film industry should be more likely to create films that are set in the states they are from. This would obviously lead to an increase in the amount of location representation in film.

Next, we have the control variable that measures the population in the state