Senior Honor Thesis Project

Preliminary Results

Rui Li

11/30/2020

1. INTRODUCTION

2. METHODS

The primary method of predicting the new coronavirus case number in this project is to compute the new case number difference between two consecutive days (Dickey and Pantula 2002) by fitting linear regression models over the case number difference and death number difference of the past 14 consecutive days, the population of each country in 2019, and the population density of each country in 2019. This paper tests on four different models and compare them to find the best performance model.

Let C_t = be a column vector with new coronavirus case number of each country at time t reversely, where $t = i, (i = 0, 1, 2 \cdots 15)$, representing i day past the current day $t = 0$. Then the response of the models is $C_0 - C_1$, and $C_i - C_{i+1}$ is the case number difference on day $i, (i = 1, 2, \cdots, 15)$. The expression of the first model is

- $C_0 - C_1 = \alpha_0 + \sum_{i=1}^{14} \alpha_i (C_i - C_{i+1}),$ (2.1)

where α_i are the estimators of i_{th} day case number difference.

Let D_t be a column vector with new coronavirus death number of each country at time t reversely, where $t = i, (i = 0, 1, 2 \cdots 15)$, representing i day past the current day $t = 0$. For the second model, we want to consider the death number difference of the past 14 consecutive days as predictors, among which $D_j - D_{j+1}$ is the death number difference on day $j, (j = 1, 2 \cdots 15)$. The expression of the second model is

- $C_0 - C_1 = \beta_0 + \sum_{i=1}^{14} \beta_i (C_i - C_{i+1}) + \sum_{j=1}^{14} \beta_{j+14} (D_j - D_{j+1}),$

where β_0 = intercept, $\beta_i, (i = 1, 2, \cdots, 14)$ = the estimators of i_{th} day case number difference, and $\beta_{j+14}, (j = 1, 2, \cdots, 14)$ = the estimators of j_{th} day death number difference.

3. RESULTS

4. DISCUSSIONS

References

Dickey, David A, and Sastry G Pantula. 2002. "Determining the Order of Differencing in Autoregressive Processes." *Journal of Business & Economic Statistics* 20:1: 18–24. https://doi.org/10.1198/073500102753410363.