October 22, 2023
Ian Verheyden

Regression Discontinuity Design/ Effect of Christmas 2020 Lockdown on daily COVID 19 case counts in Quebec

When analyzing the effect of the Christmas 2020 lockdown ("the event") on the daily COVID-19 positive case rate in Quebec, certain design considerations were necessary. First, the amount of time to consider leading up to and following the event was critical to ensure a reliable model. The window of time between November 1, 2020 and February 23rd so as to capture enough data on each side of the event to establish a reliable trend, but not so much data that the effects of other events in the timeline of the pandemic would start to influence the model. The event date was established as January 5th, 2020 (8 days after Christmas Day, the start of the lockdown) in consideration of the approximately week long incubation period for the disease.

The decision to use a linear regression model as opposed to a polynomial model was influenced by two of the arguments against the latter made by German and Imbens[1]. Namely, that polynomial models give too much influence to increasingly large values of x. In the context of our analysis of COVID cases over time, the further number of days away from the event (established as "0 days since event" in my code), the more influence the case count for that day in the model because of the squaring of the time feature. Furthermore, when polynomial models were established for the dates in question, it was seen that they gave less reliable p values for the coefficients associated with the features, when compared to the linear model.

When considering the daily Covid19 new case count for the dates between November 1st and February 23rd, we applied a linear regression model using the python statsmodels api. This model was trained on case count data as the dependent variable and the day the patient was tested as the independent variable. A binary feature was also added ("Threshold" = 0|1) depending on if the date of the positive patient test occurred before or after an event date.

When the event date of January 5th, 2021 was chosen (8 days after the beginning of the event at the centre of this section of our RDD analysis: the Dec 25 lockdown) the model calculated that before the event date, the daily case count was increasing on average about 25 cases per day (with 95% confidence of being between 23 and 28 more cases than the day before) reaching approximately 2500 cases per day, just before the lockdown. The model calculates with 95% confidence that the immediate effect of the lockdown was a reduction of between 176.5 and 508 cases per day, with the expected value of 342. The number of cases per day then decreased by an average of 59 cases per day (with 95% confidence that this number was between 53 and 64 cases per day, each day).

---

[1]   Andrew Gelman & Guido Imbens (2019) Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs, Journal of Business & Economic Statistics, 37:3, 447-456, DOI: 10.1080/07350015.2017.1366909