Quiz 3

Advance Business Analytics   25 points

Due February 16

This question tests your understanding **using time varying covariates** in the Cox model.

The Apple app store requires that all apps in the store must display a privacy label. The goal of privacy label is to inform users about app's privacy posture and let them make decision about whether to download the app. There has been a lot of interest in whether privacy labels affect demand[1] for apps. From December 14 onwards, Appstore mandated that any new or any existing app must display privacy label. I am attaching a dataset on date **when** the Apps display the label. Apps are followed for a certain number of days and if the Label appears then it is listed as 1, otherwise 0. One would expect that labels for all apps appear on or around Dec 14. However, that does not happen. Existing apps do not display privacy labels despite the Appstore mandate. This leads to additional investigation into what factors cause apps to display labels.

There are many useful attributes that may influence the timing for the labels. For example, the rank of apps on Dec 14 and type of the app (Free, paid, grossing). Appstore publishes the rank of apps based on how many downloads they receive. A higher rank (1, 2, 3.. and so on) app gets more downloads. We have reasons to believe that top ranks apps are more likely to disclose label earlier.

I am attaching a dataset to answer this question. The data provides information on the number of days it takes for an app to display the Label. The column "days_followed" lists the number of days the app was followed. The "Label" column shows whether the label appeared.

One can estimate a Cox model with the rank and app type as covariates to estimate the time it takes for app to adopt the label. However, you suspect that what other apps in a **category** do play a role in an app's incentive to adopt the label. If more apps in the focal app's category are displaying labels, the focal app may adopt the label faster. In short, you believe that the number of other apps that have adopted labels at a given time may affect the time to release the label for an app.

To test whether this has an impact on the hazard of adoption, you create a long form version of the dataset where for every focal app, for each day you count the number of apps which have released the label. Show your steps and compute this variable. [10]

For example: for a focal app in category 1 that takes 15 days to release label, you create 15 rows for the app, (one for each day). You count the number of other apps in category 1 that have released the label. That number will go up from day 1 to day 15 (it's cumulative). You can do this for each app.

Use that as a covariate in your Cox regression along with the Rank and App type and report the results, along with the model that you estimated. What is the impact of the number of other apps on the hazard? [15]

---

[1] I have been doing a project on this question.