

Homework II

[C|U]Quantile Regressions and MLE

Fernando Rios-Avila

Part I: Conditional Quantile Regressions

Consider the dataset `atus_2020_22.dta`. This dataset is a subsample of the American Time Use survey for the years 2020 through 2022. The dataset contains information on time dedicated by individuals 15 years old or older to household production activities. It also contains information on the average MET value of their activities during the day (x100) (`avgmet`). So a person with an `avgmet` value of 100 is spending the same amount of energy as if they were sitting down all day.

- Propose a model to analyze the determinants of AvgMet activities, justify the variables you include in the model.
- Estimate the model using Linear Regression, and for Conditional quantiles at 0.1, 0.25, 0.5, 0.75, and 0.9. Interpret the results of the variables you consider most important or interesting.

Part II: Unconditional Quantile Regressions

Consider the dataset `hprice2.dta`, available with `frause`. Propose a model to analyze the determinants of housing prices, making sure you also include data on `crime` and `nox` (nitrogen oxide concentration).

- Estimate unconditional quantile regressions at 0.1, 0.25, 0.5, 0.75, and 0.9. And interpret the results of the variables you consider most important.
- Say that there is a policy that promises to reduce average crime by 50%, and average NOX by 50%. How would this affect the distribution of housing prices?

Part III: Maximum Likelihood Estimation

Maximum likelihood estimation is a method to estimate the parameters of a model by maximizing the likelihood of the data given the parameters.

In this part, you will estimate the parameters of a Poisson model using maximum likelihood estimation, when the data is censored.

The Poisson Model

The Poisson model is a model for count data. The likelihood function for a single observation in a Poisson model is given by:

$$P(y = y_i) = \frac{e^{-\lambda} \lambda^{y_i}}{y_i!}$$

Where y_i is the count of events, and λ is the mean of the distribution. If we assume λ depends on a set of covariates, then we can parameterize $\lambda = \exp x\beta$.

For the purpose of programming the likelihood function, it is easier to work with the log-likelihood function, which is given by:

$$L_i(y, X\beta) = -\exp(x\beta) + y_i(x\beta) - \log(y_i!)$$

Which corresponds to line 16 in the do-file `mypoisson.do`.

Censored Data

If data is censored, the likelihood function needs to be modified to account for the fact that we do not observe the exact number of events. Specifically:

if y_i is the number of events, and c is the censoring threshold, then the likelihood function is given by:

$$\begin{aligned} \text{if } y > c &\rightarrow P(y = y_i) = \frac{e^{-\lambda} \lambda^{y_i}}{y_i!} \\ \text{if } y \leq c &\rightarrow P(y \leq c) = \sum_{k=0}^c \frac{e^{-\lambda} \lambda^k}{k!} \end{aligned}$$

Exercise

Assume that you have data (y_c) that is censored. You observe the exact number of events if more than 3 occurred. If they are less than that, they are recoded as 3.

- Modify the program `mypoissonc` in `mypoisson.do` to account for this censoring, and estimate the parameters of the model.
- Using the simulated $y, y_c, x1$ and $x2$, estimate the parameters of the model using the standard Poisson model (y), and the censored poisson model (y_c). Compare the results, and discuss the implications of the censoring on the estimation of the parameters.