Statistics 203: Introduction to Regression and Analysis of Variance

Time Series Regression

Jonathan Taylor

Today’s class

- Regression with autocorrelated errors.
- Functional data.

Autocorrelation

- In the random effects model, outcomes within groups were correlated.
- Other regression applications also have correlated outcomes (i.e. errors).
- Common examples: time series data.
- Why worry? Can lead to underestimates of SE $\rightarrow$ inflated t's $\rightarrow$ false positives.

Durbin-Watson test for autocorrelation

- In regression setting, if noise is AR(1), a simple estimate of ρ is obtained by (essentially) regressing e_t onto e_{t-1}

$$\widehat{\rho} = \frac{\sum_{t=2}^{n} \left(e_t e_{t-1}\right)}{\sum_{t=1}^{n} e_t^2}.$$

- To formally test $H_0 : \rho = 0$ (i.e. whether residuals are independent vs. they are AR(1)), use Durbin-Watson test, based on

$$d = 2(1 - \widehat{\rho}).$$

Correcting for AR(1) in regression model

- If we now ρ, it is possible "pre-whiten" the data and regressors

$$\tilde{Y}_{i+1} = Y_{i+1} - \rho Y_i, i > 1$$

$$\tilde{X}_{(i+1)j} = X_{(i+1)j} - \rho X_{ij}, i > 1$$

then model satisfies "usual" assumptions.

- For coefficients in new model $\tilde{\beta}$, $\beta_0 = \tilde{\beta}_0/(1-\rho)$, $\beta_j = \tilde{\beta}_j$.

Two-stage regression

- Step 1: Fit linear model to unwhitened data.
- Step 2: Estimate ρ with $\widehat{\rho}$.
- Step 3: Pre-whiten data using $\widehat{\rho}$ – refit the model.

Other models of correlation

- If we have $ARMA(p,q)$ noise then we can also pre-whiten the data and perform OLS – equivalent to GLS.
- If we estimate parameters we can then use a two-stage procedure as in the AR(1) case.
- OR, we can just use MLE (or REML): R does this. This is similar to iterating the two-stage procedure.

More than one time series

- Suppose we have r time series $Y_{ij}, 1 \leq i \leq r, 1 \leq j \leq n_r$.
- Regression model

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \varepsilon_{ij}.$$

where the β's are common to everyone and

$$\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i}) \sim N(0, \Sigma_i),$$

independent across i

- We can put all of this into one big regression model and estimate *everything.* Easy to do in R.

Functional Data

- Having observations that are time series can be thought of as having a "function" as an observation.
- Having many time series, i.e. daily temperature in NY, SF, LA, . . . allows one to think of the individual time series as observations.
- The field "Functional Data Analysis" (Ramsay & Silverman) is a part of statistics that focuses on this type of data.
- Today we'll think of having one function and what we might do with it.

Scatterplot smoothing

- When we only have one "function" we can think of fitting a trend as smoothing a scatterplot of pairs $(X_i, Y_i)_{1 \leq i \leq n}$.
- Different techniques
  - B-splines;
  - Smoothing splines;
  - Kernel smoothers;
  - many others.

Smoothing splines

- We saw early on in the class that we could use B-splines in a regression setting to predict Y_i from X_i.
- Smoothing splines: for $\lambda \geq 0$ and weights $w_i, 1 \leq i \leq n$ find the function with two-derivatives that minimizes

$$\sum_{i=1}^{n} \omega_i (Y_i - f(X_i))^2 + \lambda \int (f''(x))^2 \, dx.$$

- This should remind you of ridge regression: prior is now on *functions*.
- Equivalent to saying that we have a Gaussian prior (integrated Brownian motion) on functions and we want the "MAP" estimator based on observing f at the points X with measurement errors $\varepsilon_i \sim N(0, 1/w_i)$.

Kernel smoother

- Given a kernel function K and a bandwidth h, the kernel smooth of the scatterplot $(X_i, Y_i)_{1 \leq i \leq n}$ is defined by the local average

$$\widehat{Y}(x) = \frac{\sum_{i=1}^n Y_i \cdot K((x - X_i)/h)}{\sum_{i=1}^n K((x - X_i)/h)}.$$

- Most commonly used kernel:

$$K(x) = e^{-x^2/2}.$$

- The key parameter is the bandwidth. Much work has been done on choosing an "optimal bandwidth."