Lab Session 1: Linear regression with t errors

Hedibert Freitas Lopes

The University of Chicago Booth School of Business
5807 South Woodlawn Avenue, Chicago, IL 60637
http://faculty.chicagobooth.edu/hedibert.lopes

hlopes@ChicagoBooth.edu

NLSY Data

We revisit the NLSY data where y_i is hourly wage (in logs) received by individual i and x_i is his/her years of schooling completed, i.e. (for $i = 1, \ldots, n$)

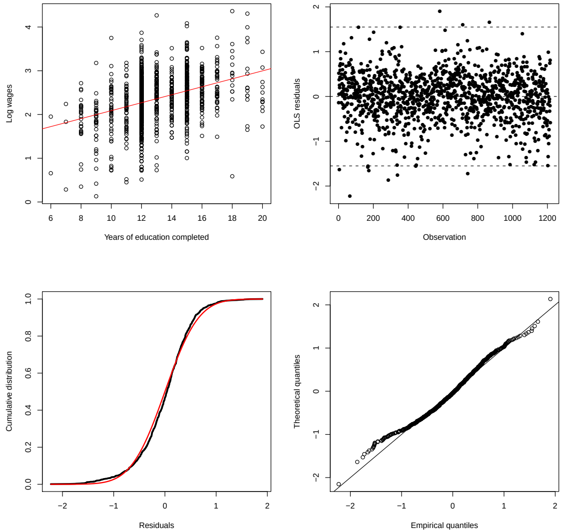
$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i.$$

Competing models:
Gaussian model ($\mathcal{M}_0$): $\epsilon_i \sim N(0, \sigma^2)$
Student's t model ($\mathcal{M}_1$): $\epsilon_i \sim t_\nu(0, \sigma^2)$

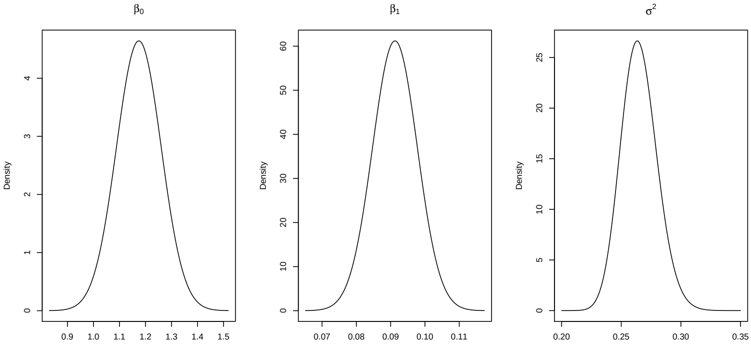
Prior distribution:
$\beta | \sigma^2 \sim N(0, \sigma^2 10 I_2)$
$\sigma^2 \sim IG(3, 2.5)$
$\nu \sim G(1, 25)$ (Geweke, 1993).

This prior specification allocates substantial prior probability on values of ν below 10 (fat-tails) as well as above 40 (normality).

OLS regression



Normal linear regression model



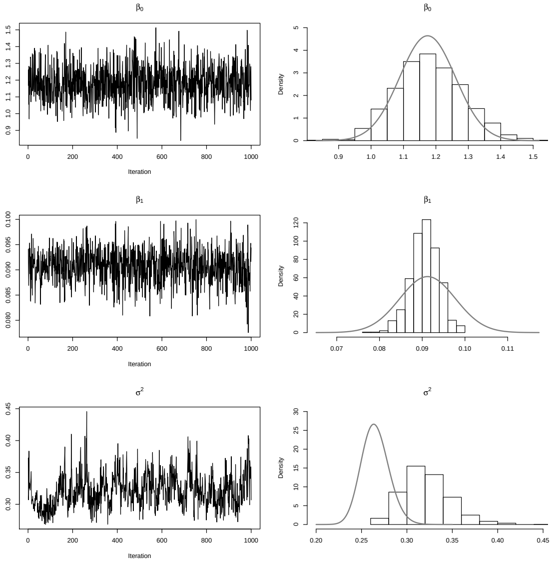
$\hat{\beta} = (1.178, 0.091)$ and $\hat{\sigma}^2 = 0.267$.
$E(\beta|y, x) = (1.174, 0.091)$ and $E(\sigma^2|y, x) = 0.265$.

Posterior inference

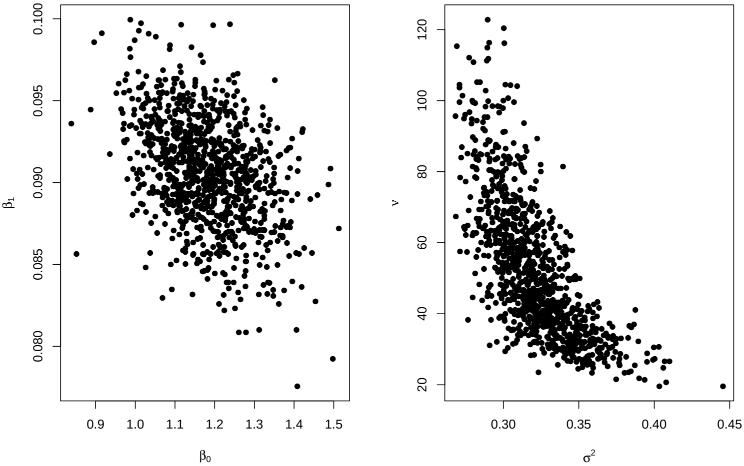
We run a MCMC algorithm with starting at the the OLS estimates for β_0, β_1 and σ^2 and at $\nu^{(0)} = 1$.

The chains were warmed-up for a $M_0 = 1000$ iterations and every 100-th of the following 100,000 draws were kept for posterior summarization, producing a total of $M = 1000$ draws.

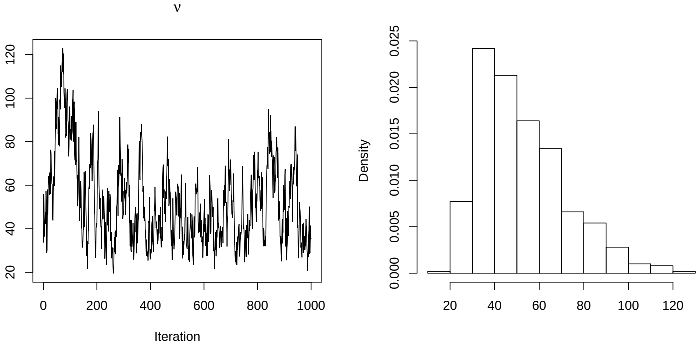
MCMC output


β0
β0
β1
β1
σ2
σ2
Iteration
Iteration
Iteration
Density
Density
Density

Bivariate marginal posterior



Normal versus Student's t



95% posterior credibility interval for ν is $(27, 100)$.
33% is the prior probability for the same interval.