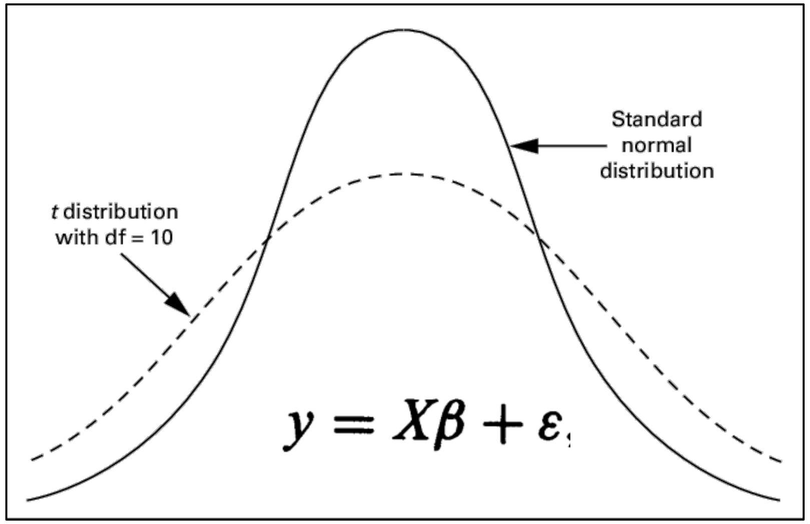
Aspects of Robust Regression Analysis

Supervisors: Nancy Reid & Archer Gong Zhang
Research Students: Amanda Ng, Shangkai Zhu

Introduction

- Limitations of traditional regression models (e.g. OLS):
  - Outliers
  - Efficiency when error is not normal
  - High dimension
- Proposed solution: Student-T regression models
  - Frequentist (MLE)
  - Bayesian (MAP)
- Parameters of student-t likelihood: β, σ, ν



Frequentist approach using Profile likelihood

1. Likelihood of Student-T distribution:

$$L(\beta, \sigma, \nu; y, x) = \frac{\Gamma(\frac{\nu+1}{2})^n \nu^{n\nu/2}}{\Gamma(\frac{\nu}{2})^n \pi^{n/2} \sigma^n} \prod_{i=1}^{n} \left\{ \nu + (\frac{y_i - x_i^\top \beta}{\sigma})^2 \right\}^{-(\nu+1)/2}$$

2. Nuisance parameter $\lambda = (\beta, \sigma)$

Optimize

3. Constrained MLE $(\hat{\beta}_\nu, \hat{\sigma}_\nu) = \arg\max L(\beta, \sigma, \nu; y, x)$

4. Profile likelihood $L^p(\hat{\beta}_\nu, \hat{\sigma}_\nu, \nu; y, x) = L(\hat{\beta}_\nu, \hat{\sigma}_\nu, \nu; y, x)$

Optimize

5. MLE

Frequentist approach using Adjusted profile likelihood

1. Adjusted profile log likelihood

$$\ell_{\mathrm{adj}}(\nu) = \ell_{\mathrm{p}}(\nu) - \frac{1}{2}\log|j_{\lambda\lambda}(\nu, \hat{\lambda}_\nu)|,^{*}$$

Optimize

2. Adjusted MLE

$$\hat{\nu}_{\mathrm{adj}} := \arg\max_{\nu} \ell_{\mathrm{adj}}(\nu),$$

*Observed Fisher information nuisance parameters block

$$j_{\lambda\lambda}(\nu,\lambda) = \begin{pmatrix} -\frac{\partial^2 \ell}{\partial^2 \beta} & -\frac{\partial^2 \ell}{\partial \beta \partial \sigma} \\ -\frac{\partial^2 \ell}{\partial \beta \partial \sigma} & -\frac{\partial^2 \ell}{\partial^2 \sigma} \end{pmatrix} = \begin{pmatrix} \frac{\nu+1}{\sigma^2}\sum_{i=1}^n \frac{x_i x_i^\top}{\nu+z_i^2} - \frac{2z_i^2 x_i x_i^\top}{(\nu+z_i^2)^2} & \frac{(\nu+1)2\nu}{\sigma^3}\sum_{i=1}^n \frac{(y_i - x_i^\top \beta)x_i}{\left(\nu+\left(\frac{y_i - x_i^\top\beta}{\sigma}\right)^2\right)^2} \\ \frac{(\nu+1)2\nu}{\sigma^3}\sum_{i=1}^n \frac{(y_i - x_i^\top\beta)x_i^\top}{(\nu+z_i^2)^2} & -\frac{n}{\sigma^2} + (\nu+1)\sum_{i=1}^n \frac{3}{\sigma^4}\frac{(y_i - x_i^\top\beta)^2}{\nu+z_i^2} - \frac{2}{\sigma^6}\frac{(y_i - x_i^\top\beta)^4}{(\nu+z_i^2)^2} \end{pmatrix}$$

Bayesian approach using priors

1. Improper priors

$$\begin{cases} \beta \propto 1 & \text{Flat prior} \\ \sigma \propto 1/\sigma & \text{Inverse prior} \\ \nu \propto \left(\frac{\nu}{\nu+3}\right)^{1/2} \left\{\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right\}^{1/2} & \text{Jeffrey's prior} \end{cases}$$

2. Profile likelihood with constrained MLE

3. Posterior distribution $\xrightarrow{\text{Optimize}}$ 4. MAP

Evaluate performance

Simulate covariate data (X) with using fixed true β, true σ

Loop over a range of true ν

Simulate Y (e.g. 50 repetitions) using true ν and X

For each Y, optimize for:

- MLE
- Adjusted MLE
- MAP

**Exclude those that do not converge successfully*

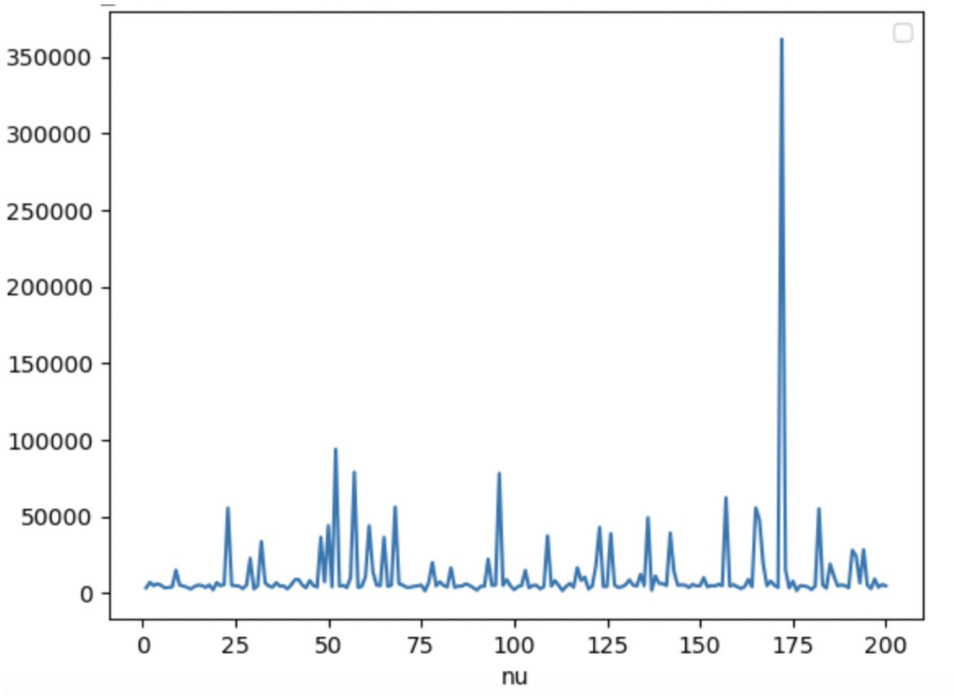
Calculate averaged MSE for each approach

Plot MSE vs true ν

Challenges

- Convergence issue in optimizations

Log profile likelihood

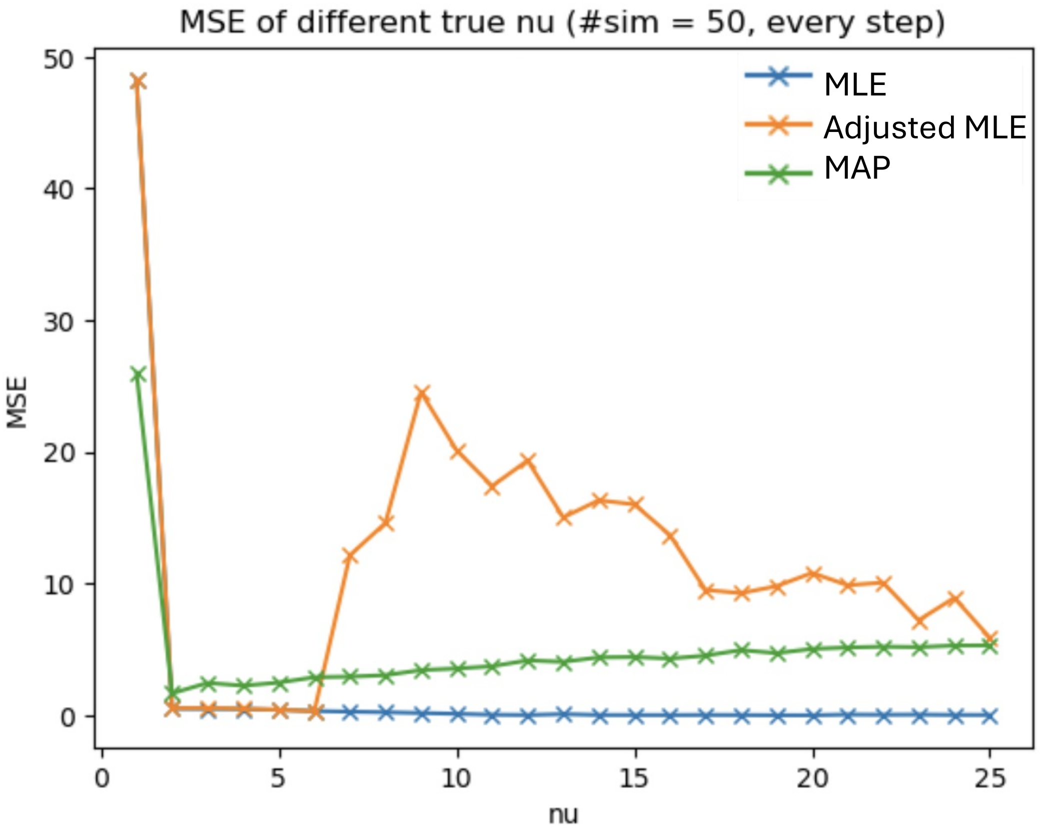


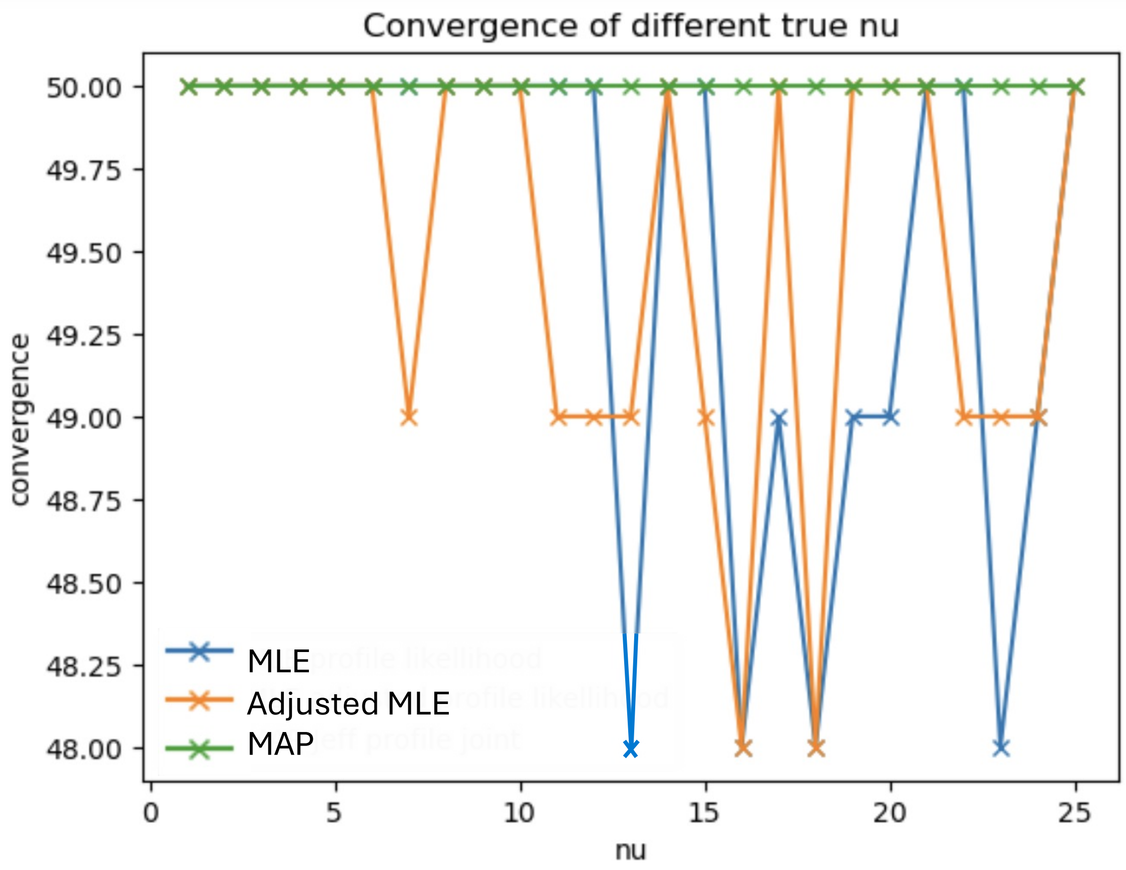
- Run Time issue in two-step optimizations and PyMC MAP function

Solutions

1. Try a combination of optimization **algorithms**
    - *BFGS* for first step (to get constrained β, σ MLEs)
    - *Nelder-Mead* for second step (to get ν MLE)
2. Adjust **initial guess** in optimizations
    - Least Square estimates for β, σ
    - True ν
3. Address the problem of nu being **non-positive**
    - Take $\log \nu$ as input for all functions
4. Optimize **all parameters** simultaneously on **Bayesian joint density**
    - Skip constrained MLE

Results about ν





Conclusion

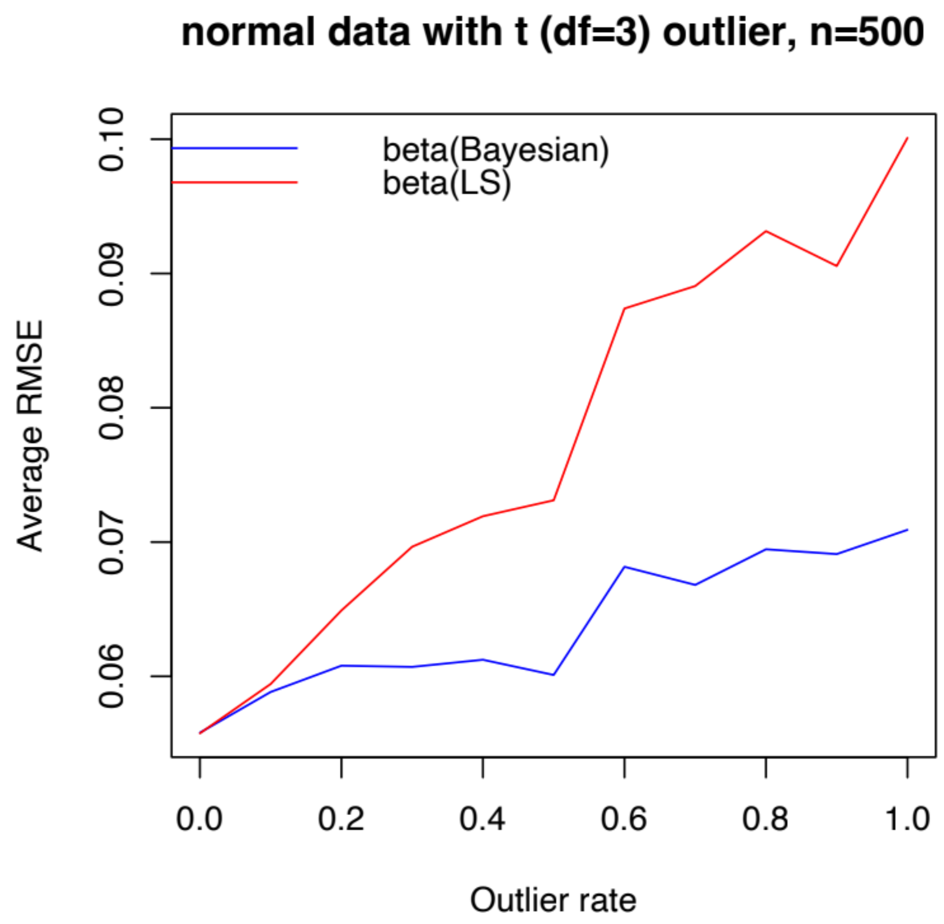
- Good estimation of ν is hard
- T-regression should be more robust than ordinary least squares

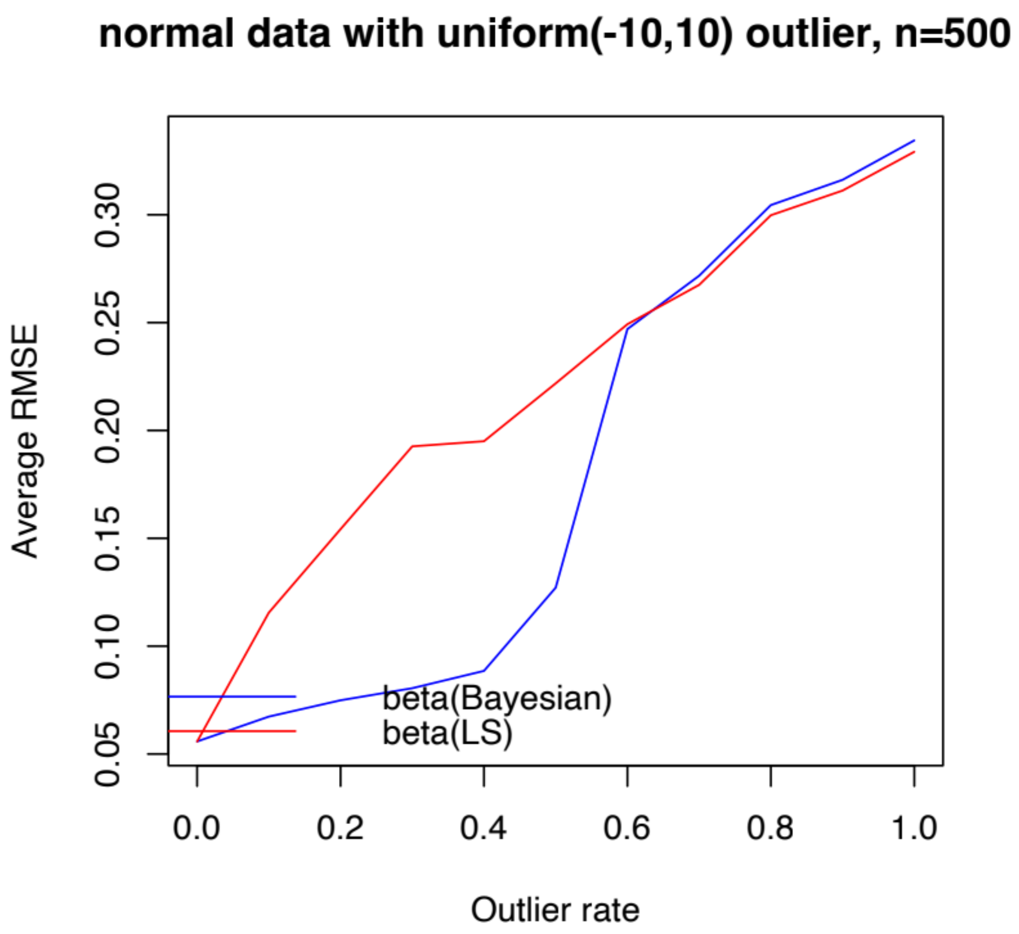
Future Directions

- Investigate the coverage of confidence intervals for regression parameters β
- Study the robustness of inference with T-regression compared to least squares

Supplementary Slides

Results about outliers on β





Results about β confidence interval