Valid standard errors for misspecified Bayesian models

Sophia Rabe-Hesketh

Education & Biostatistics
University of California, Berkeley
sophiarh@berkeley.edu

Joint work with Feng Ji and JoonHo Lee

Stata Conference, July 31, 2025

Reminiscence on occasion of Stata's 40th anniversay

- 5th UK Stata User Group Meeting, 1999, my first talk on GLLAMM
- Amazing support, led to **`gllamm`** being sped up by Stata developers!

Reminiscence on occasion of Stata's 40th anniversay

- 5th UK Stata User Group Meeting, 1999, my first talk on GLLAMM
- Amazing support, led to **`gllamm`** being sped up by Stata developers!
- First five Stata User Group Meetings were all in London
  - At Royal Statistical Society, Errol Street
  - We used transparencies!

Reminiscence on occasion of Stata's 40th anniversay

- 5th UK Stata User Group Meeting, 1999, my first talk on GLLAMM
- Amazing support, led to **`gllamm`** being sped up by Stata developers!
- First five Stata User Group Meetings were all in London
  - At Royal Statistical Society, Errol Street
  - We used transparencies!
  - Bill Gould always present
  - Wishes & Grumbles sessions

Outline

1. Bayesian Infinitesimal Jacknife (IJ) standard errors (SEs)
2. Standard Bayesian quantile regression is misspecified
3. IJ SEs for Bayesian quantile regression
4. IJ SEs for clusterd data and functions of parameters
5. Discussion

1. Bayesian IJ SEs

Standard Bayesian estimation

- Assume model $p(D|\theta)$ for data D with parameter vector $\theta = (\theta_1, \ldots, \theta_p)'$; $p(D|\theta)$ is the likelihood

Standard Bayesian estimation

- Assume model $p(D|\theta)$ for data D with parameter vector $\theta = (\theta_1, \ldots, \theta_p)'$; $p(D|\theta)$ is the likelihood
- Specify prior $p(\theta)$ for parameters

Standard Bayesian estimation

- Assume model $p(D|\theta)$ for data D with parameter vector $\theta = (\theta_1, \ldots, \theta_p)'$; $p(D|\theta)$ is the likelihood
- Specify prior $p(\theta)$ for parameters
- Posterior $p(\theta|D) \propto p(\theta)p(D|\theta)$ used for Bayesian inference

Standard Bayesian estimation

- Assume model $p(D|\theta)$ for data D with parameter vector $\theta = (\theta_1, \ldots, \theta_p)'$; $p(D|\theta)$ is the likelihood
- Specify prior $p(\theta)$ for parameters
- Posterior $p(\theta|D) \propto p(\theta)p(D|\theta)$ used for Bayesian inference
- Posterior expectation, $E(\theta_r \mid D)$, is point estimator of parameter θ_r
  - in MCMC, approximated by average of S posterior samples, $\tilde{\theta}_r = \frac{1}{S}\sum_{s=1}^{S} \theta_r^{(s)}$

Standard Bayesian estimation

- Assume model $p(D|\theta)$ for data D with parameter vector $\theta = (\theta_1, \ldots, \theta_p)'$; $p(D|\theta)$ is the likelihood
- Specify prior $p(\theta)$ for parameters
- Posterior $p(\theta|D) \propto p(\theta)p(D|\theta)$ used for Bayesian inference
- Posterior expectation, $E(\theta_r \mid D)$, is point estimator of parameter θ_r
  - in MCMC, approximated by average of S posterior samples, $\tilde{\theta}_r = \frac{1}{S}\sum_{s=1}^{S} \theta_r^{(s)}$
- Posterior standard deviation $sd(\theta_r \mid D)$ expresses uncertainty of belief about θ_r given this dataset D
  - in MCMC, approximated by standard deviation of posterior samples, $s_r = \sqrt{\frac{1}{S-1}\sum_{s=1}^{S}(\theta_r^{(s)} - \tilde{\theta}_r)^2}$

Use frequentist SEs when likelihood is misspecified

- Frequentist SE is standard deviation of Bayesian point estimates in repeated samples

Use frequentist SEs when likelihood is misspecified

- Frequentist SE is standard deviation of Bayesian point estimates in repeated samples
- If likelihood $p(D|\theta)$ is correct [Bernstein-Von Mises]
  - Point estimator is consistent
  - $\mathrm{sd}(\theta_r \mid D)$ and frequentist SE coincide asymptotically
  - Credible intervals and frequentist CIs coincide asymptotically

Use frequentist SEs when likelihood is misspecified

- Frequentist SE is standard deviation of Bayesian point estimates in repeated samples
- If likelihood $p(D|\theta)$ is correct [Bernstein-Von Mises]
  - Point estimator is consistent
  - $\mathrm{sd}(\theta_r \mid D)$ and frequentist SE coincide asymptotically
  - Credible intervals and frequentist CIs coincide asymptotically
- If likelihood is misspecified
  - Point estimator converges to pseudo-true parameter
    [Kleijn & van der Vaart (2012)]

Use frequentist SEs when likelihood is misspecified

- Frequentist SE is standard deviation of Bayesian point estimates in repeated samples
- If likelihood $p(D|\theta)$ is correct [Bernstein-Von Mises]
  - Point estimator is consistent
  - $\mathrm{sd}(\theta_r \mid D)$ and frequentist SE coincide asymptotically
  - Credible intervals and frequentist CIs coincide asymptotically
- If likelihood is misspecified
  - Point estimator converges to pseudo-true parameter [Kleijn & van der Vaart (2012)]
  - $p(\theta|D)$ and $\mathrm{sd}(\theta_r \mid D)$ not correct for Bayesian inference

Use frequentist SEs when likelihood is misspecified

- Frequentist SE is standard deviation of Bayesian point estimates in repeated samples
- If likelihood $p(D|\theta)$ is correct [Bernstein-Von Mises]
  - Point estimator is consistent
  - $\mathrm{sd}(\theta_r \mid D)$ and frequentist SE coincide asymptotically
  - Credible intervals and frequentist CIs coincide asymptotically
- If likelihood is misspecified
  - Point estimator converges to pseudo-true parameter [Kleijn & van der Vaart (2012)]
  - $p(\theta|D)$ and $\mathrm{sd}(\theta_r \mid D)$ not correct for Bayesian inference
  - Frequentist SE can be meaningful

Use frequentist SEs when likelihood is misspecified

- Frequentist SE is standard deviation of Bayesian point estimates in repeated samples
- If likelihood $p(D|\theta)$ is correct [Bernstein-Von Mises]
  - Point estimator is consistent
  - $\mathrm{sd}(\theta_r \mid D)$ and frequentist SE coincide asymptotically
  - Credible intervals and frequentist CIs coincide asymptotically
- If likelihood is misspecified
  - Point estimator converges to pseudo-true parameter [Kleijn & van der Vaart (2012)]
  - $p(\theta|D)$ and $\mathrm{sd}(\theta_r \mid D)$ not correct for Bayesian inference
  - Frequentist SE can be meaningful
- Methods for obtaining (asymptotic) frequentist SEs:
  - Sandwich estimator: By Integration/Laplace, not based on MCMC

Use frequentist SEs when likelihood is misspecified

- Frequentist SE is standard deviation of Bayesian point estimates in repeated samples
- If likelihood $p(D|\theta)$ is correct [Bernstein-Von Mises]
  - Point estimator is consistent
  - $\mathrm{sd}(\theta_r \mid D)$ and frequentist SE coincide asymptotically
  - Credible intervals and frequentist CIs coincide asymptotically
- If likelihood is misspecified
  - Point estimator converges to pseudo-true parameter [Kleijn & van der Vaart (2012)]
  - $p(\theta|D)$ and $\mathrm{sd}(\theta_r \mid D)$ not correct for Bayesian inference
  - Frequentist SE can be meaningful
- Methods for obtaining (asymptotic) frequentist SEs:
  - Sandwich estimator: By Integration/Laplace, not based on MCMC
  - Nonparametric bootstrapping: Time-consuming to perform MCMC in many bootstrap samples

Use frequentist SEs when likelihood is misspecified

- Frequentist SE is standard deviation of Bayesian point estimates in repeated samples
- If likelihood $p(D|\theta)$ is correct [Bernstein-Von Mises]
  - Point estimator is consistent
  - $\mathrm{sd}(\theta_r \mid D)$ and frequentist SE coincide asymptotically
  - Credible intervals and frequentist CIs coincide asymptotically
- If likelihood is misspecified
  - Point estimator converges to pseudo-true parameter [Kleijn & van der Vaart (2012)]
  - $p(\theta|D)$ and $\mathrm{sd}(\theta_r \mid D)$ not correct for Bayesian inference
  - Frequentist SE can be meaningful
- Methods for obtaining (asymptotic) frequentist SEs:
  - Sandwich estimator: By Integration/Laplace, not based on MCMC
  - Nonparametric bootstrapping: Time-consuming to perform MCMC in many bootstrap samples
  - IJ SEs: Computed from one MCMC run!

Infinitesimal Jacknife SEs [Giordano & Broderick, 2024]

- Start with idea of resampling (e.g., Jackknife or bootstrap)

Infinitesimal Jacknife SEs [Giordano & Broderick, 2024]

- Start with idea of resampling (e.g., Jackknife or bootstrap)
  - Weight-vector w with elements w_i, $i = 1, \ldots, n$, representing how many times unit i was sampled

Infinitesimal Jacknife SEs [Giordano & Broderick, 2024]

- Start with idea of resampling (e.g., Jackknife or bootstrap)
  - Weight-vector w with elements w_i, $i = 1, \ldots, n$, representing how many times unit i was sampled
  - Log-likelihood for resampled data: $\sum_{i=1}^{n} w_i \ell_i(D|\theta)$
    - ◇ $\ell_i(D|\theta)$ is log-likelihood contribution from unit i

Infinitesimal Jacknife SEs [Giordano & Broderick, 2024]

- Start with idea of resampling (e.g., Jackknife or bootstrap)
  - Weight-vector w with elements w_i, $i = 1, \ldots, n$, representing how many times unit i was sampled
  - Log-likelihood for resampled data: $\sum_{i=1}^{n} w_i \ell_i(D|\theta)$
    - ◇ $\ell_i(D|\theta)$ is log-likelihood contribution from unit i
- Linear approximation of Bayesian estimator in resampled data

$$E(\theta \mid D, w) \approx \underbrace{E(\theta \mid D, w = 1_n)}_{\text{for actual data}} + \left.\frac{dE(\theta \mid D, w)}{dw'}\right|_{w=1_n} (w - 1_n)$$

Infinitesimal Jacknife SEs [Giordano & Broderick, 2024]

- Start with idea of resampling (e.g., Jackknife or bootstrap)
  - Weight-vector w with elements w_i, $i = 1, \ldots, n$, representing how many times unit i was sampled
  - Log-likelihood for resampled data: $\sum_{i=1}^{n} w_i \ell_i(D|\theta)$
    - $\ell_i(D|\theta)$ is log-likelihood contribution from unit i
- Linear approximation of Bayesian estimator in resampled data

$$E(\theta \mid D, w) \approx \underbrace{E(\theta \mid D, w = 1_n)}_{\text{for actual data}} + \left.\frac{dE(\theta \mid D, w)}{dw'}\right|_{w=1_n} (w - 1_n)$$

$$= E(\theta \mid D, w = 1_n) + \mathrm{cov}_{\theta|D}[\theta, \ell(D \mid \theta)](w - 1_n)$$

  - $(p \times n)$ posterior covariance of θ and $\ell(D|\theta)$, vector of $\ell_i(D|\theta)$, estimated by empirical covariance matrix of MCMC samples

Infinitesimal Jacknife SEs [Giordano & Broderick, 2024]

- Start with idea of resampling (e.g., Jackknife or bootstrap)
  - Weight-vector w with elements w_i, $i = 1, \ldots, n$, representing how many times unit i was sampled
  - Log-likelihood for resampled data: $\sum_{i=1}^{n} w_i \ell_i(D|\theta)$
    - $\ell_i(D|\theta)$ is log-likelihood contribution from unit i
- Linear approximation of Bayesian estimator in resampled data

$$E(\theta \mid D, w) \approx \underbrace{E(\theta \mid D, w = 1_n)}_{\text{for actual data}} + \left. \frac{dE(\theta \mid D, w)}{dw'} \right|_{w=1_n} (w - 1_n)$$

$$= E(\theta \mid D, w = 1_n) + \text{cov}_{\theta|D}[\theta, \ell(D \mid \theta)](w - 1_n)$$

  - $(p \times n)$ posterior covariance of θ and $\ell(D|\theta)$, vector of $\ell_i(D|\theta)$, estimated by empirical covariance matrix of MCMC samples
- Influence score $I_i := n\, \text{cov}_{\theta|D}[\theta, \ell_i(D \mid \theta)]$

Infinitesimal Jacknife SEs [Giordano & Broderick, 2024]

- Start with idea of resampling (e.g., Jackknife or bootstrap)
  - Weight-vector w with elements w_i, $i = 1, \ldots, n$, representing how many times unit i was sampled
  - Log-likelihood for resampled data: $\sum_{i=1}^{n} w_i \ell_i(D|\theta)$
    - ⋄ $\ell_i(D|\theta)$ is log-likelihood contribution from unit i
- Linear approximation of Bayesian estimator in resampled data

$$E(\theta \mid D, w) \approx \underbrace{E(\theta \mid D, w = 1_n)}_{\text{for actual data}} + \left.\frac{dE(\theta \mid D, w)}{dw'}\right|_{w=1_n} (w - 1_n)$$

$$= E(\theta \mid D, w = 1_n) + \mathrm{cov}_{\theta|D}[\theta, \ell(D \mid \theta)](w - 1_n)$$

  - $(p \times n)$ posterior covariance of θ and $\ell(D|\theta)$, vector of $\ell_i(D|\theta)$, estimated by empirical covariance matrix of MCMC samples
- Influence score $I_i := n\,\mathrm{cov}_{\theta|D}[\theta, \ell_i(D \mid \theta)]$
- IJ Variance (squared IJ SEs on diagonal) based on MCMC estimates $\hat{I}_i$

$$\hat{V}^{\mathsf{IJ}} := \frac{1}{n(n-1)} \sum_{i=1}^{n} (\hat{I}_i - \overline{\hat{I}})(\hat{I}_i - \overline{\hat{I}})'$$

2. Standard Bayesian quantile regression is misspecified

Classical quantile regression [Koenker & Bassett, 1978]

- Linear regression is model for $E(y|x)$ as a function of covariates x
  $E(y|x) = x'\beta$

Classical quantile regression [Koenker & Bassett, 1978]

- Linear regression is model for $E(y|x)$ as a function of covariates x
  $E(y|x) = x'\beta$
- Quantile regression is a model for conditional quantiles
  $Q_\tau(y|x) = x'\beta(\tau)$
  - τ is quantile level, e.g. $\tau = 0.5$ gives median regression
  - Makes no assumption regarding conditional distribution of y given x

Classical quantile regression [Koenker & Bassett, 1978]

- Linear regression is model for $E(y|x)$ as a function of covariates x

  $E(y|x) = x'\beta$

- Quantile regression is a model for conditional quantiles

  $Q_\tau(y|x) = x'\beta(\tau)$

  - τ is quantile level, e.g. $\tau = 0.5$ gives median regression
  - Makes no assumption regarding conditional distribution of y given x

- Frequentist estimator minimizes a loss function:

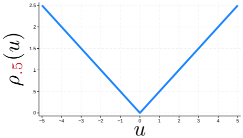
  $\hat{\beta}(\tau) = \text{argmin}_{\beta(\tau)} \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta(\tau))$

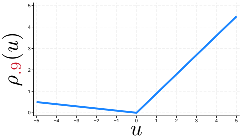
Classical quantile regression [Koenker & Bassett, 1978]

- Linear regression is model for $E(y|x)$ as a function of covariates x
  $E(y|x) = x'\beta$
- Quantile regression is a model for conditional quantiles
  $Q_\tau(y|x) = x'\beta(\tau)$
  - τ is quantile level, e.g. $\tau = 0.5$ gives median regression
  - Makes no assumption regarding conditional distribution of y given x
- Frequentist estimator minimizes a loss function:

  $\hat{\beta}(\tau) = \text{argmin}_{\beta(\tau)} \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta(\tau))$

  - $\rho_\tau(u) = u\{\tau - I(u < 0)\} = \begin{cases} u\tau & \text{if } u \geq 0 \\ -u(1-\tau) & \text{if } u < 0 \end{cases}$





Standard Bayesian quantile regression [Yu & Moyeed, (2001)]

- Need a likelihood!
  Choose exponential of minus scaled classical loss function

  $p(D|\theta) \propto \exp\{-\sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta(\tau))/\sigma\}$

  - Produces Gibbs posterior distribution [Syring & Martin (2019)]

Standard Bayesian quantile regression [Yu & Moyeed, (2001)]

- Need a likelihood!
  Choose exponential of minus scaled classical loss function

  $$p(D|\theta) \propto \mathsf{exp}\{-\sum_{i=1}^{n} \rho_\tau(\underbrace{y_i - x_i'\beta(\tau)}_{\epsilon_i})/\sigma\}$$

  - Produces Gibbs posterior distribution [Syring & Martin (2019)]
- Corresponds to asymmetric Laplace (AL) density for $\epsilon_i|x_i$

  $$y_i = x_i'\beta(\tau) + \epsilon_i$$

  $$f_{\mathsf{AL}}(\epsilon_i|\theta, x_i) = \frac{\tau(1-\tau)}{\sigma}\mathsf{exp}\left\{-\rho_\tau\left(\frac{\epsilon_i}{\sigma}\right)\right\}$$

Standard Bayesian quantile regression [Yu & Moyeed, (2001)]

- Need a likelihood!
  Choose exponential of minus scaled classical loss function

  $$p(D|\theta) \propto \exp\{-\sum_{i=1}^{n} \rho_\tau(\underbrace{y_i - x_i'\beta(\tau)}_{\epsilon_i})/\sigma\}$$

  - Produces Gibbs posterior distribution [Syring & Martin (2019)]
- Corresponds to asymmetric Laplace (AL) density for $\epsilon_i|x_i$

  $$y_i = x_i'\beta(\tau) + \epsilon_i$$
  $$f_{\mathsf{AL}}(\epsilon_i|\theta, x_i) = \frac{\tau(1-\tau)}{\sigma}\exp\left\{-\rho_\tau\left(\frac{\epsilon_i}{\sigma}\right)\right\}$$

- Analogy: Likelihood based on exponential of minus scaled sum of squared errors corresponds to normal density

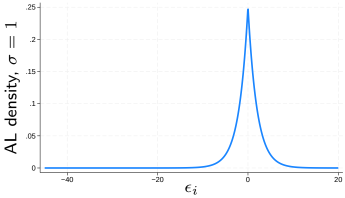
AL likelihood is misspecified

- "Working likelihood," chosen because MLE is classical estimator

AL likelihood is misspecified

- "Working likelihood," chosen because MLE is classical estimator
- Highly restrictive, implausible as data-generating model

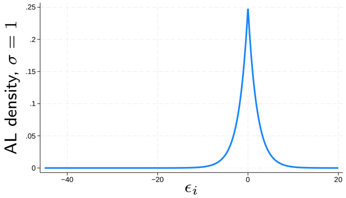
$\tau = 0.5$, AL is symmetric, SD is 2.8



AL likelihood is misspecified

- "Working likelihood," chosen because MLE is classical estimator
- Highly restrictive, implausible as data-generating model

$\tau = 0.5$, AL is symmetric, SD is 2.8

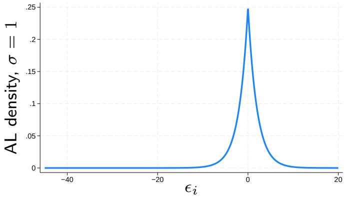


  - Assumes homoscedasticity, i.e., parallel quantiles!

AL likelihood is misspecified

- "Working likelihood," chosen because MLE is classical estimator
- Highly restrictive, implausible as data-generating model

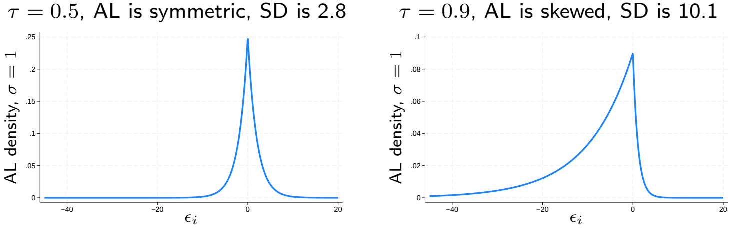
$\tau = 0.5$, AL is symmetric, SD is 2.8



  - Assumes homoscedasticity, i.e., parallel quantiles!
  - Assumes specific spacing of quantiles!

AL likelihood is misspecified

- "Working likelihood," chosen because MLE is classical estimator
- Highly restrictive, implausible as data-generating model

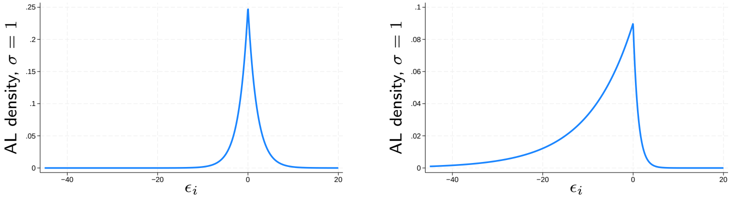


  - Assumes homoscedasticity, i.e., parallel quantiles!
  - Assumes specific spacing of quantiles!
  - Distribution changes, depending on value of τ we are interested in

AL likelihood is misspecified

- "Working likelihood," chosen because MLE is classical estimator
- Highly restrictive, implausible as data-generating model

$\tau = 0.5$, AL is symmetric, SD is 2.8 $\qquad$ $\tau = 0.9$, AL is skewed, SD is 10.1



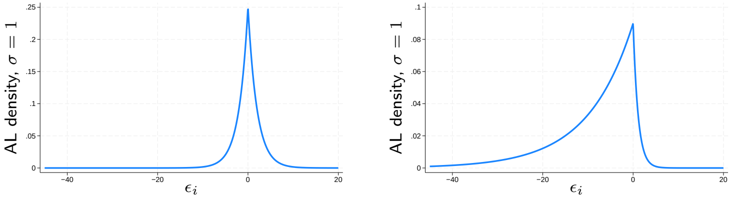
  - Assumes homoscedasticity, i.e., parallel quantiles!
  - Assumes specific spacing of quantiles!
  - Distribution changes, depending on value of τ we are interested in

- $\Rightarrow$ Cannot trust $sd(\theta_r \mid D)$

AL likelihood is misspecified

- "Working likelihood," chosen because MLE is classical estimator
- Highly restrictive, implausible as data-generating model

$\tau = 0.5$, AL is symmetric, SD is 2.8 $\tau = 0.9$, AL is skewed, SD is 10.1



  - Assumes homoscedasticity, i.e., parallel quantiles!
  - Assumes specific spacing of quantiles!
  - Distribution changes, depending on value of τ we are interested in

- $\Rightarrow$ Cannot trust $sd(\theta_r \mid D)$
- Asymptotically, $sd(\theta_r \mid D)$ proportional to $\sqrt{\sigma}$
[Sriram, 2015; Yang et al., (2016)]
  - But scale parameter σ for working likelihood seems arbitrary

Approaches to AL-based Bayesian quantile regression

- Use $sd(\theta_r \mid D)$ to quantify uncertainty

Approaches to AL-based Bayesian quantile regression

- Use $sd(\theta_r \mid D)$ to quantify uncertainty
  - ☹ Set $\sigma = 1$ [e.g., Yu & Moseed (2001)]
    $sd(\theta_r \mid D)$ meaningless!
    As bad as setting $\sigma = 1$ in linear regression

Approaches to AL-based Bayesian quantile regression

- Use $sd(\theta_r \mid D)$ to quantify uncertainty
  - Set $\sigma = 1$ [e.g., Yu & Moseed (2001)]
    $sd(\theta_r \mid D)$ meaningless!
    As bad as setting $\sigma = 1$ in linear regression
  - Specify prior for σ
    Treats AL as correct error distribution!
    - **bayes:qreg** in Stata and **BayesQR** in R: inverse Gamma
    - **brms** in R: half-$t(3)$

Approaches to AL-based Bayesian quantile regression

- Use $sd(\theta_r \mid D)$ to quantify uncertainty
  - ☹ Set $\sigma = 1$ [e.g., Yu & Moseed (2001)]
    $sd(\theta_r \mid D)$ meaningless!
    As bad as setting $\sigma = 1$ in linear regression
  - 😐 Specify prior for σ
    Treats AL as correct error distribution!
    - **`bayes:qreg`** in Stata and **`BayesQR`** in R: inverse Gamma
    - **`brms`** in R: half-$t(3)$
- Disregard $sd(\theta_r \mid D)$ [Yang et al. (2016); Sriram (2015); Lee (2020); Ji (2022)]
  - Adjusted SE [Yang et al. (2016)] based on asymptotic SE of MLE
    - sets σ to a constant
    - **`AdjBQR`** in R sets σ to MLE at $\tau = 0.5$
  - Sandwich likelihood [Sriram (2015)]

Approaches to AL-based Bayesian quantile regression

- Use $sd(\theta_r \mid D)$ to quantify uncertainty
  - Set $\sigma = 1$ [e.g., Yu & Moseed (2001)]
    $sd(\theta_r \mid D)$ meaningless!
    As bad as setting $\sigma = 1$ in linear regression
  - Specify prior for σ
    Treats AL as correct error distribution!
    - **bayes:qreg** in Stata and **BayesQR** in R: inverse Gamma
    - **brms** in R: half-$t(3)$
- Disregard $sd(\theta_r \mid D)$ [Yang et al. (2016); Sriram (2015); Lee (2020); Ji (2022)]
  - Adjusted SE [Yang et al. (2016)] based on asymptotic SE of MLE
    - sets σ to a constant
    - **AdjBQR** in R sets σ to MLE at $\tau = 0.5$
  - Sandwich likelihood [Sriram (2015)]
  - IJ SEs [Ji, Lee & Rabe-Hesketh (2025)]

3. IJ SEs for Bayesian quantile regression

Simulation study

- Model

  $y_i = \alpha + \beta x_i + (1 + \gamma x_i)\epsilon_i, \quad \epsilon_i | x_i \sim N(0, 1)$

  $\Rightarrow Q_\tau(y_i \mid x_i) = [\alpha + \Phi^{-1}(\tau)] + [\beta + \gamma\Phi^{-1}(\tau)]x_i$

Simulation study

- Model

  $y_i = \alpha + \beta x_i + (1 + \gamma x_i)\epsilon_i, \quad \epsilon_i | x_i \sim N(0,1)$

  $\Rightarrow Q_\tau(y_i \mid x_i) = [\alpha + \Phi^{-1}(\tau)] + [\beta + \gamma\Phi^{-1}(\tau)]x_i$

- Conditions: Fix $\alpha = \beta = 2$, $\gamma = 0.3$ and vary τ, σ, and n

Simulation study

- Model

  $y_i = \alpha + \beta x_i + (1 + \gamma x_i)\epsilon_i, \quad \epsilon_i | x_i \sim N(0,1)$

  $\Rightarrow Q_\tau(y_i \mid x_i) = [\alpha + \Phi^{-1}(\tau)] + [\beta + \gamma\Phi^{-1}(\tau)]x_i$

- Conditions: Fix $\alpha = \beta = 2$, $\gamma = 0.3$ and vary τ, σ, and n

- Methods
  - Frequentist: **boot**, **sandwich**
  - Proposed here: **IJ** with σ estimated and **IJf** with σ fixed arbitrarily
  - Adjusted [Yang et al., 2016)]:
    - **Yang** with σ fixed arbitrarily
    - **AdjBQR** with σ set to MLE at $\tau = 0.5$
  - Bayesian
    - **brms** with half-$t(3)$ prior for σ
    - **BayesQR** with inverse Gamma(.01, .01) for σ

Simulation study

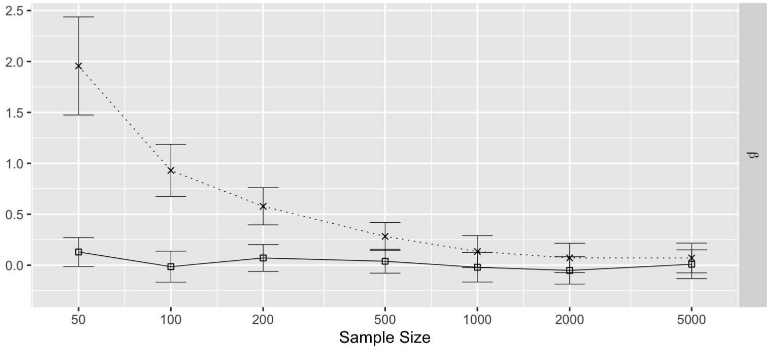
- Model

  $y_i = \alpha + \beta x_i + (1 + \gamma x_i)\epsilon_i, \quad \epsilon_i | x_i \sim N(0, 1)$

  $\Rightarrow Q_\tau(y_i \mid x_i) = [\alpha + \Phi^{-1}(\tau)] + [\beta + \gamma\Phi^{-1}(\tau)]x_i$
- Conditions: Fix $\alpha = \beta = 2$, $\gamma = 0.3$ and vary τ, σ, and n
- Methods
  - Frequentist: **boot**, **sandwich**
  - Proposed here: **IJ** with σ estimated and **IJf** with σ fixed arbitrarily
  - Adjusted [Yang et al., 2016)]:
    - ◇ **Yang** with σ fixed arbitrarily
    - ◇ **`AdjBQR`** with σ set to MLE at $\tau = 0.5$
  - Bayesian
    - ◇ **`brms`** with half-$t(3)$ prior for σ
    - ◇ **`BayesQR`** with inverse Gamma(.01, .01) for σ
- Evaluate Relative error (with 95% CI [White (2010)])
  - $R_e = \sqrt{\frac{\overline{\text{se}^2}}{\text{var}(\widehat{\beta})}} - 1,$
    - ◇ $\overline{\text{se}^2}$ is average squared SE, $\text{var}(\widehat{\beta})$ is variance of estimate

Relative error with fixed, large $\sigma = 20$

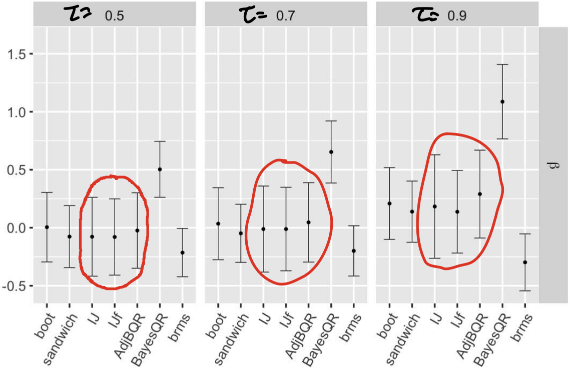
- $\sigma = 20$, $\tau = 0.7$, increasing n



-  **IJf** performs well even for small n
-  **Yang** requires larger n to perform well

Relative error with σ estimated or fixed at $\sigma = 1$

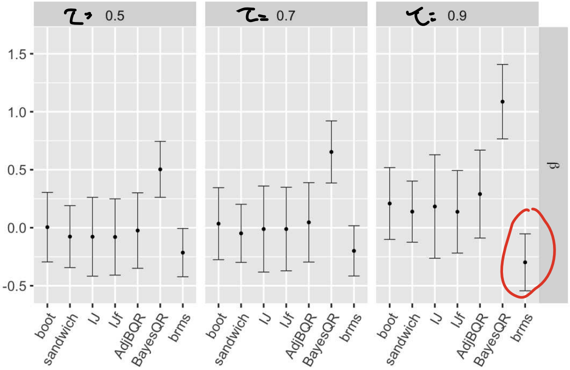
- $n = 200$, vary τ



- frequentist, **IJ**, **IJf** and **AdjBQR** perform well and similarly

Relative error with σ estimated or fixed at $\sigma = 1$

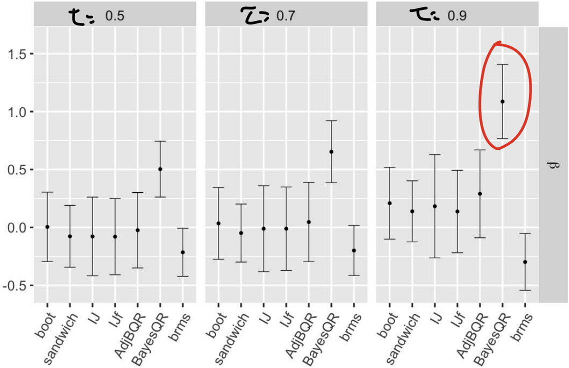
- $n = 200$, vary τ



- frequentist, **IJ**, **IJf** and **AdjBQR** perform well and similarly
- **brms** underestimates SE at $\tau = 0.9$

Relative error with σ estimated or fixed at $\sigma = 1$

- $n = 200$, vary τ



- frequentist, **IJ**, **IJf** and **AdjBQR** perform well and similarly
- **brms** underestimates SE at $\tau = 0.9$
- **BayesQR** greatly overestimates SE, by over 75% at $\tau = 0.9$

Credible/confidence intervals for **engel1857.dta**

- Engel's (1857) hypothesis:

  "The poorer a family, the greater the part of total expenditures must be spent on food"

Credible/confidence intervals for `engel1857.dta`

- Engel's (1857) hypothesis:

  "The poorer a family, the greater the part of total expenditures must be spent on food"

- Subjects: 235 European working-class households

Credible/confidence intervals for **engel1857.dta**

- Engel's (1857) hypothesis:

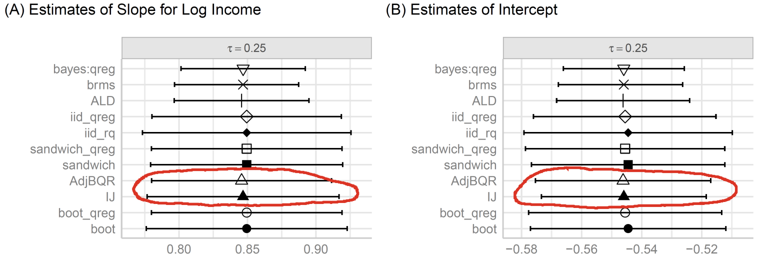
  "The poorer a family, the greater the part of total expenditures must be spent on food"
- Subjects: 235 European working-class households
- Analysis: Quantile regression of log food expenditure on log income to estimate "Engel elasticities" [Koenker & Bassett (1982)]

Credible/confidence intervals for `engel1857.dta`

- Engel's (1857) hypothesis:

  "The poorer a family, the greater the part of total expenditures must be spent on food"

- Subjects: 235 European working-class households
- Analysis: Quantile regression of log food expenditure on log income to estimate "Engel elasticities" [Koenker & Bassett (1982)]



- **IJ** and **`AdjBQR`** CIs similar to frequentist CIs

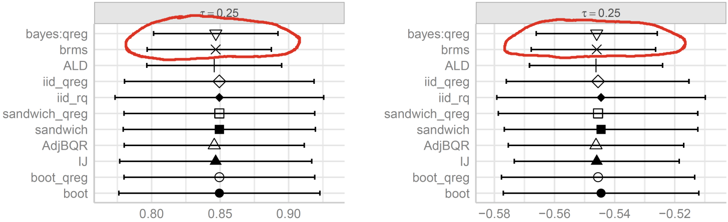
Credible/confidence intervals for **engel1857.dta**

- Engel's (1857) hypothesis:

  "The poorer a family, the greater the part of total expenditures must be spent on food"
- Subjects: 235 European working-class households
- Analysis: Quantile regression of log food expenditure on log income to estimate "Engel elasticities" [Koenker & Bassett (1982)]

(A) Estimates of Slope for Log Income

(B) Estimates of Intercept



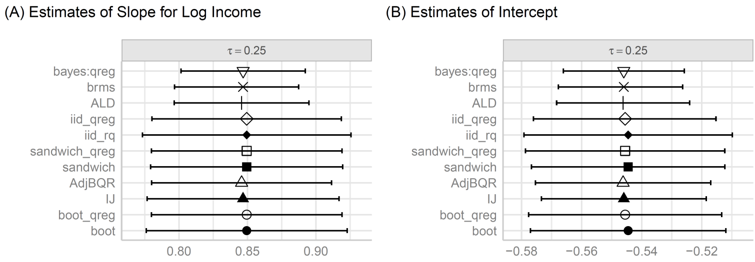
- **IJ** and **AdjBQR** CIs similar to frequentist CIs
- **bayes:qreg** and **brms** CIs too narrow

Credible/confidence intervals for `engel1857.dta`

- Engel's (1857) hypothesis:

  "The poorer a family, the greater the part of total expenditures must be spent on food"

- Subjects: 235 European working-class households
- Analysis: Quantile regression of log food expenditure on log income to estimate "Engel elasticities" [Koenker & Bassett (1982)]



- **`IJ`** and **`AdjBQR`** CIs similar to frequentist CIs
- **`bayes:qreg`** and **`brms`** CIs too narrow
- **`BayesQR`** badly off and therefore omitted

4. IJ SEs for clustered data and functions of parameters

Influence scores for clusters

- Define influence score $I_j^{(cl)}$ for cluster j, $j = 1, \ldots, J$ (motivate by resampling clusters)
  - Starting with influence scores for units $I_i := n\,\mathrm{cov}_{\theta|D}[\theta, \ell_i(D \mid \theta)]$, influence score for cluster is

$$I_j^{(cl)} := \frac{J}{n} \sum_{\substack{i \\ \text{in cluster } j}} I_i$$

Influence scores for clusters

- Define influence score $I_j^{(cl)}$ for cluster j, $j = 1, \ldots, J$ (motivate by resampling clusters)
  - Starting with influence scores for units $I_i := n\,\mathrm{cov}_{\theta|D}[\theta, \ell_i(D \mid \theta)]$, influence score for cluster is

$$I_j^{(cl)} := \frac{J}{n} \sum_{\substack{i \\ \text{in cluster } j}} I_i$$

  - Equivalently, starting with cluster log-likelihood contributions

$$\ell_j^{(cl)} := \sum_{\substack{i \\ \text{in cluster } j}} \ell_i(D \mid \theta),$$

influence score for cluster is $I_j^{(cl)} := J\,\mathrm{cov}_{\theta|D}\left[\theta, {\ell_j}^{(cl)}(D \mid \theta)\right]$

IJ SEs for clustered data

- Estimate $\hat{I}_j^{(cl)}$ from MCMC samples
- IJ variance is

$$\hat{V}_{(cl)}^{\mathsf{IJ}} := \frac{1}{J(J-1)} \sum_{j=1}^{J} (\hat{I}_j^{(cl)} - \overline{\hat{I}^{(cl)}})(\hat{I}_j^{(cl)} - \overline{\hat{I}^{(cl)}})'$$

Functions of parameters

- Vector of functions of parameters $f(\theta)$
  - Indirect effect in linear mediation is product of coefficients
  - Reliability in measurement is ratio of variance parameters
  - etc.
- Influence score for IJ variance becomes

$$I_i := n\,\mathrm{cov}_{\theta|D}[f(\theta), \ell_i(D \mid \theta)]$$

5. Discussion

AL-based Bayesian quantile regression

- Naïve posterior standard deviations continue to be used (**`brms`**, **`bayes:qreg`**, many papers)

AL-based Bayesian quantile regression

- Naïve posterior standard deviations continue to be used (**`brms`**, **`bayes:qreg`**, many papers)
- Adjusted SEs [Yang et al. (2015)] work well if σ estimated by MLE at $\tau = 0.5$, as in **`AdjBQR`**

AL-based Bayesian quantile regression

- Naïve posterior standard deviations continue to be used (**`brms`**, **`bayes:qreg`**, many papers)
- Adjusted SEs [Yang et al. (2015)] work well if σ estimated by MLE at $\tau = 0.5$, as in **`AdjBQR`**
- But IJ SEs preferable because they work for:
  - general σ and small sample sizes
  - clustered data & functions of parameters
  - other models!

AL-based Bayesian quantile regression

- Naïve posterior standard deviations continue to be used (**`brms`**, **`bayes:qreg`**, many papers)
- Adjusted SEs [Yang et al. (2015)] work well if σ estimated by MLE at $\tau = 0.5$, as in **`AdjBQR`**
- But IJ SEs preferable because they work for:
  - general σ and small sample sizes
  - clustered data & functions of parameters
  - other models!
- Comment on point estimates of $\beta(\tau)$ [Ji, Lee & Rabe-Hesketh (2025)]
  - Posterior becomes more skewed as σ increases for $\tau \neq 0.5$, leading to posterior means larger (smaller) than posterior mode/MLE for $\tau > 0.5$ ($\tau < 0.5$)
  - Decrease σ if posterior skewed

Other advantages of IJ SEs

- Applicable for any Bayesian model
  - Assumptions often doubtful, e.g., homoscedasticity
  - Clustered data common
  - Potential to become as popular in Bayesian setting as sandwich estimator in frequentist setting!

Some "Wishes and Grumbles"

- Wish: Make IJ SEs available for all Bayesian models
  - Add option to **`bayesmh`** and **`bayes`** prefix command?
  - Introduce **`bayesstats IJSE`**?

Some "Wishes and Grumbles"

- Wish: Make IJ SEs available for all Bayesian models
  - Add option to **bayesmh** and **bayes** prefix command?
  - Introduce **bayesstats IJSE**?
- Wish/Grumble: Acknowledge misspecification of AL likelihood in **bayes:qreg**
  - Explain in documentation
  - Provide warning in output and provide IJ SEs by default
  - Disable (or provide warning for) **sigma()** option
  - Disable model-based postestimation
    e.g., **bayesstats ic**, **bayesstest model**, **bayespredict**

Some "Wishes and Grumbles"

- Wish: Make IJ SEs available for all Bayesian models
  - Add option to **`bayesmh`** and **`bayes`** prefix command?
  - Introduce **`bayesstats IJSE`**?
- Wish/Grumble: Acknowledge misspecification of AL likelihood in **`bayes:qreg`**
  - Explain in documentation
  - Provide warning in output and provide IJ SEs by default
  - Disable (or provide warning for) **`sigma()`** option
  - Disable model-based postestimation
    e.g., **`bayesstats ic`**, **`bayesstest model`**, **`bayespredict`**

Thank You!

References related to quantile regression

- Hagemann, A. (2017). Cluster-robust bootstrap inference in quantile regression models. *Journal of the American Statistical Association*, *112*, 446–456.
- Ji, F. (2022). *Practically feasible solutions to a set of problems in applied statistics*. Doctoral dissertation, University of California, Berkeley.
- Ji, F., Lee, J.-H. & Rabe-Hesketh, S. (2025). Valid standard errors for Bayesian quantile regression with clustered and independent data. *Journal of Educational and Behavioral Statistics*, conditionally accepted.
- Koenker, R., & Bassett, G. S. (1978). Regression quantiles. *Econometrica*, *46*, 33–50.
- Koenker, R., & Bassett, G. S. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, 50, 43–61.
- Lee, J.-H. (2020). *Essays on treatment effect heterogeneity in education policy interventions*. Doctoral dissertation, University of California, Berkeley.
- Sriram, K. (2015). A sandwich likelihood correction for Bayesian quantile regression based on the misspecified asymmetric Laplace density. *Statistics & Probability Letters*, *107*, 18–26.
- Yang, Y., Wang, H. J., & He, X. (2016). Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood. *International Statistical Review*, *84*, 327–344.
- Yu, K., & Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, *54*, 437–447.

Other references

- Giordano, R., & Broderick, T. (2024). The Bayesian infinitesimal jackknife for variance. *arXiv preprint arXiv:2305.06466*
- Kleijn, B. and A. van der Vaart (2012). The Bernstein-von-Mises theorem under misspecification. *Electronic Journal of Statistics* 6, 354-381.
- Syring, N., & Martin, R. (2019). Calibrating general posterior credible regions. *Biometrika*, *106*, 479–486.
- White, I. R. (2010). simsum: Analyses of simulation studies including Monte Carlo error. *The Stata Journal*, *10*, 369–375.