Bayesian Quantile Mixture Regression

Athanasios Kottas

Department of Statistics, University of California, Santa Cruz

Joint work with Yifei Yan (*Amazon, New York*)

12th International Conference on Bayesian Nonparametrics
Oxford, U.K., June 24-28, 2019

Quantile regression

- Quantile regression quantifies the relationship between a set of quantiles of the response distribution and covariates, thus providing a more complete explanation of the response distribution.

- Practically important alternative to traditional mean regression models, with a growing literature in terms of methods and applications.

- Single-quantile regression formulation: $y_i = h(\boldsymbol{x}_i) + \varepsilon_i$, $i = 1, \ldots, n$
  - Response observations y_i, with covariate vectors $\boldsymbol{x}_i$.
  - ε_i i.i.d. (given parameters) from an error distribution with density $f_p(\varepsilon)$ and p-th quantile equal to 0, i.e., $\int_{-\infty}^{0} f_p(\varepsilon)\mathrm{d}\varepsilon = p$.
  - $h(\boldsymbol{x})$ is the quantile regression function, e.g., $h(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta}$, for a linear quantile regression model.

Quantile regression

- Quantile regression quantifies the relationship between a set of quantiles of the response distribution and covariates, thus providing a more complete explanation of the response distribution.

- Practically important alternative to traditional mean regression models, with a growing literature in terms of methods and applications.

- Single-quantile regression formulation: $y_i = h(\boldsymbol{x}_i) + \varepsilon_i$, $i = 1, \ldots, n$
  - Response observations y_i, with covariate vectors $\boldsymbol{x}_i$.
  - ε_i i.i.d. (given parameters) from an error distribution with density $f_p(\varepsilon)$ and p-th quantile equal to 0, i.e., $\int_{-\infty}^{0} f_p(\varepsilon)\mathrm{d}\varepsilon = p$.
  - $h(\boldsymbol{x})$ is the quantile regression function, e.g., $h(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta}$, for a linear quantile regression model.

Estimation and modeling methods

- **Classical nonparametric estimation:** point estimates for quantile regression coefficients $\boldsymbol{\beta}$ through optimization of the *check loss function*, $\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})$, where $\rho_p(u) = up - u1_{(u<0)}$ (Koenker, 2005).
  - Least absolute deviations criterion, $\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} |y_i - \boldsymbol{x}_i'\boldsymbol{\beta}|$, for $p = 0.5$.
- **Bayesian modeling and inference:**
  - Nonparametric priors for the error distribution (Kottas & Gelfand, 2001; Hanson & Johnson, 2002; Kottas & Krnjajić, 2009; Reich et al., 2010)
  - Joint estimation of multiple quantile regression curves (Tokdar & Kadane, 2012; Reich & Smith, 2013; Yang & Tokdar, 2017)
  - Nonparametric inference through density regression (Taddy & Kottas, 2010)
  - Parametric modeling based on the asymmetric Laplace (AL) distribution (Yu & Moyeed, 2001; Tsionas, 2003; Kozumi & Kobayashi, 2011)

Estimation and modeling methods

- **Classical nonparametric estimation:** point estimates for quantile regression coefficients $\boldsymbol{\beta}$ through optimization of the *check loss function*, $\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})$, where $\rho_p(u) = up - u1_{(u<0)}$ (Koenker, 2005).
  - Least absolute deviations criterion, $\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} |y_i - \boldsymbol{x}_i'\boldsymbol{\beta}|$, for $p = 0.5$.

- **Bayesian modeling and inference:**
  - Nonparametric priors for the error distribution (Kottas & Gelfand, 2001; Hanson & Johnson, 2002; Kottas & Krnjajić, 2009; Reich et al., 2010)
  - Joint estimation of multiple quantile regression curves (Tokdar & Kadane, 2012; Reich & Smith, 2013; Yang & Tokdar, 2017)
  - Nonparametric inference through density regression (Taddy & Kottas, 2010)
  - Parametric modeling based on the asymmetric Laplace (AL) distribution (Yu & Moyeed, 2001; Tsionas, 2003; Kozumi & Kobayashi, 2011)

Objectives

- **Quantile mixture regression:** modeling for the response distribution through weighted mixtures of quantile regression components.
  - Common regression function (common β in linear regression) for all mixture components.
  - Synthesize information from multiple parts of the response distribution for estimation and selection of covariate effects.
  - Different from simultaneous quantile regression, the goal is to obtain a combined estimate of the predictive effect of each covariate.
- Mixtures of K distributions, each parameterized in terms of the p_k-th quantile, for $k = 1, ..., K$.
  - Fixed p_k: mixtures of generalized AL distributions (more flexible skewness and tail behavior than the AL, when p_k is fixed).
  - Random p_k: mixtures of AL distributions.

Objectives

- **Quantile mixture regression:** modeling for the response distribution through weighted mixtures of quantile regression components.
  - Common regression function (common $\boldsymbol{\beta}$ in linear regression) for all mixture components.
  - Synthesize information from multiple parts of the response distribution for estimation and selection of covariate effects.
  - Different from simultaneous quantile regression, the goal is to obtain a combined estimate of the predictive effect of each covariate.

- Mixtures of K distributions, each parameterized in terms of the p_k-th quantile, for $k = 1, ..., K$.
  - Fixed p_k: mixtures of generalized AL distributions (more flexible skewness and tail behavior than the AL, when p_k is fixed).
  - Random p_k: mixtures of AL distributions.

Objectives

- **Quantile mixture regression:** modeling for the response distribution through weighted mixtures of quantile regression components.
  - Common regression function (common $\boldsymbol{\beta}$ in linear regression) for all mixture components.
  - Synthesize information from multiple parts of the response distribution for estimation and selection of covariate effects.
  - Different from simultaneous quantile regression, the goal is to obtain a combined estimate of the predictive effect of each covariate.
- Mixtures of K distributions, each parameterized in terms of the p_k-th quantile, for $k = 1, ..., K$.
  - Fixed p_k: mixtures of generalized AL distributions (more flexible skewness and tail behavior than the AL, when p_k is fixed).
  - Random p_k: mixtures of AL distributions.

Objectives

- **Quantile mixture regression:** modeling for the response distribution through weighted mixtures of quantile regression components.
  - Common regression function (common $\boldsymbol{\beta}$ in linear regression) for all mixture components.
  - Synthesize information from multiple parts of the response distribution for estimation and selection of covariate effects.
  - Different from simultaneous quantile regression, the goal is to obtain a combined estimate of the predictive effect of each covariate.
- Mixtures of K distributions, each parameterized in terms of the p_k-th quantile, for $k = 1, \ldots, K$.
  - Fixed p_k: mixtures of generalized AL distributions (more flexible skewness and tail behavior than the AL, when p_k is fixed).
  - Random p_k: mixtures of AL distributions.

Modeling framework

- Quantile mixture regression model:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \theta)$$

  - Continuous response y, covariate vector $\boldsymbol{x}$.
  - The p_k-th quantile for density f_{p_k} is given by $\mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}$ (or, more generally, by $\mu_{p_k} + h(\boldsymbol{x})$).
  - $0 < p_1 < \ldots < p_K < 1$ are (possibly random) probabilities associated with quantiles $\mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}$.
  - Restriction $\mu_{p_1} < \ldots < \mu_{p_K}$ yields the ordering constraint for the quantiles $\rightarrow$ facilitates identifiability $\rightarrow$ connects mixture weights with quantiles.
  - Additional scale/skewness/tail parameters θ (may be component specific).

Modeling framework

- Quantile mixture regression model:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \theta)$$

  - Continuous response y, covariate vector $\boldsymbol{x}$.
  - The p_k-th quantile for density f_{p_k} is given by $\mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}$ (or, more generally, by $\mu_{p_k} + h(\boldsymbol{x})$).
  - $0 < p_1 < \ldots < p_K < 1$ are (possibly random) probabilities associated with quantiles $\mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}$.
  - Restriction $\mu_{p_1} < \ldots < \mu_{p_K}$ yields the ordering constraint for the quantiles $\rightarrow$ facilitates identifiability $\rightarrow$ connects mixture weights with quantiles.
  - Additional scale/skewness/tail parameters θ (may be component specific).

Modeling framework

- Quantile mixture regression model:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \theta)$$

  - Continuous response y, covariate vector $\boldsymbol{x}$.
  - The p_k-th quantile for density f_{p_k} is given by $\mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}$ (or, more generally, by $\mu_{p_k} + h(\boldsymbol{x})$).
  - $0 < p_1 < \ldots < p_K < 1$ are (possibly random) probabilities associated with quantiles $\mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}$.
  - Restriction $\mu_{p_1} < \ldots < \mu_{p_K}$ yields the ordering constraint for the quantiles $\rightarrow$ facilitates identifiability $\rightarrow$ connects mixture weights with quantiles.
  - Additional scale/skewness/tail parameters θ (may be component specific).

Modeling framework

- Quantile mixture regression model:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \theta)$$

  - Continuous response y, covariate vector $\boldsymbol{x}$.
  - The p_k-th quantile for density f_{p_k} is given by $\mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}$ (or, more generally, by $\mu_{p_k} + h(\boldsymbol{x})$).
  - $0 < p_1 < \ldots < p_K < 1$ are (possibly random) probabilities associated with quantiles $\mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}$.
  - Restriction $\mu_{p_1} < \ldots < \mu_{p_K}$ yields the ordering constraint for the quantiles $\rightarrow$ facilitates identifiability $\rightarrow$ connects mixture weights with quantiles.
  - Additional scale/skewness/tail parameters θ (may be component specific).

Modeling framework

- Quantile mixture regression model:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \theta)$$

  - Continuous response y, covariate vector $\boldsymbol{x}$.
  - The p_k-th quantile for density f_{p_k} is given by $\mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}$ (or, more generally, by $\mu_{p_k} + h(\boldsymbol{x})$).
  - $0 < p_1 < \ldots < p_K < 1$ are (possibly random) probabilities associated with quantiles $\mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}$.
  - Restriction $\mu_{p_1} < \ldots < \mu_{p_K}$ yields the ordering constraint for the quantiles $\rightarrow$ facilitates identifiability $\rightarrow$ connects mixture weights with quantiles.
  - Additional scale/skewness/tail parameters θ (may be component specific).

Connection with quantile regression methods

- Different from simultaneous quantile regression which jointly models different covariate effects for different quantiles ($\boldsymbol{\beta}_{p_k}$ instead of $\boldsymbol{\beta}$).

- Similar in spirit with composite quantile regression from the classical literature (Zou & Yuan, 2008): for a collection of K specified quantile levels, $0 < p_1 < \ldots < p_K < 1$, the CQR estimator:

$$(\hat{b}_1, \ldots, \hat{b}_K, \hat{\boldsymbol{\beta}}) = \underset{b_1,\ldots,b_k,\boldsymbol{\beta}}{\arg\min} \sum_{k=1}^{K} \left\{ \sum_{i=1}^{n} \rho_{p_k}(y_i - b_k - \boldsymbol{x}_i'\boldsymbol{\beta}) \right\}$$

  - $b_k \equiv b_{p_k}$, for $k = 1, \ldots, K$: intercepts for the specified quantile levels.
  - Variable selection method with properties that include consistency in selection and asymptotic normality, as $n \to \infty$.
  - But, only an optimization algorithm ... does not involve/correspond to a probabilistic model.

Connection with quantile regression methods

- Different from simultaneous quantile regression which jointly models different covariate effects for different quantiles ($\boldsymbol{\beta}_{p_k}$ instead of $\boldsymbol{\beta}$).

- Similar in spirit with composite quantile regression from the classical literature (Zou & Yuan, 2008): for a collection of K specified quantile levels, $0 < p_1 < \ldots < p_K < 1$, the CQR estimator:

$$(\hat{b}_1, \ldots, \hat{b}_K, \hat{\boldsymbol{\beta}}) = \underset{b_1,\ldots,b_k,\boldsymbol{\beta}}{\arg\min} \sum_{k=1}^{K} \left\{ \sum_{i=1}^{n} \rho_{p_k}(y_i - b_k - \boldsymbol{x}_i'\boldsymbol{\beta}) \right\}$$

  - $b_k \equiv b_{p_k}$, for $k = 1, \ldots, K$: intercepts for the specified quantile levels.
  - Variable selection method with properties that include consistency in selection and asymptotic normality, as $n \to \infty$.
  - But, only an optimization algorithm ... does not involve/correspond to a probabilistic model.

Connection with quantile regression methods

- Different from simultaneous quantile regression which jointly models different covariate effects for different quantiles ($\boldsymbol{\beta}_{p_k}$ instead of $\boldsymbol{\beta}$).

- Similar in spirit with composite quantile regression from the classical literature (Zou & Yuan, 2008): for a collection of K specified quantile levels, $0 < p_1 < \ldots < p_K < 1$, the CQR estimator:

$$(\hat{b}_1, \ldots, \hat{b}_K, \hat{\boldsymbol{\beta}}) = \underset{b_1,\ldots,b_k,\boldsymbol{\beta}}{\arg\min} \sum_{k=1}^{K} \left\{ \sum_{i=1}^{n} \rho_{p_k}(y_i - b_k - \boldsymbol{x}_i'\boldsymbol{\beta}) \right\}$$

  - $b_k \equiv b_{p_k}$, for $k = 1, ..., K$: intercepts for the specified quantile levels.
  - Variable selection method with properties that include consistency in selection and asymptotic normality, as $n \to \infty$.
  - But, only an optimization algorithm ... does not involve/correspond to a probabilistic model.

Connection with quantile regression methods

- Different from simultaneous quantile regression which jointly models different covariate effects for different quantiles ($\boldsymbol{\beta}_{p_k}$ instead of $\boldsymbol{\beta}$).

- Similar in spirit with composite quantile regression from the classical literature (Zou & Yuan, 2008): for a collection of K specified quantile levels, $0 < p_1 < \ldots < p_K < 1$, the CQR estimator:

$$(\hat{b}_1, \ldots, \hat{b}_K, \hat{\boldsymbol{\beta}}) = \underset{b_1,\ldots,b_k,\boldsymbol{\beta}}{\arg\min} \sum_{k=1}^{K} \left\{ \sum_{i=1}^{n} \rho_{p_k}(y_i - b_k - \boldsymbol{x}_i'\boldsymbol{\beta}) \right\}$$

  - $b_k \equiv b_{p_k}$, for $k = 1, \ldots, K$: intercepts for the specified quantile levels.
  - Variable selection method with properties that include consistency in selection and asymptotic normality, as $n \to \infty$.
  - But, only an optimization algorithm ... does not involve/correspond to a probabilistic model.

Two modeling scenarios

- We envision two modeling scenarios (with fixed K throughout):
  - fixed p_k: for settings where one expects specific parts of the response distribution (e.g., the right tail) to be informed by the covariates;
  - the fixed p_k model can also be used when such information is not available, based on a set of equally spaced p_k spanning the unit interval;
  - random p_k: specify only the total number of quantiles, and let the data inform the full configuration of quantile components (both p_k and μ_{p_k}).
- Mixture components need to be flexible distributions parameterized in terms of quantiles.
  - For the random p_k model, the AL distribution is sufficiently flexible.
  - But not so when the quantile level is fixed $\rightarrow$ generalized AL distribution.

Two modeling scenarios

- We envision two modeling scenarios (with fixed K throughout):
  - fixed p_k: for settings where one expects specific parts of the response distribution (e.g., the right tail) to be informed by the covariates;
  - the fixed p_k model can also be used when such information is not available, based on a set of equally spaced p_k spanning the unit interval;
  - random p_k: specify only the total number of quantiles, and let the data inform the full configuration of quantile components (both p_k and μ_{p_k}).
- Mixture components need to be flexible distributions parameterized in terms of quantiles.
  - For the random p_k model, the AL distribution is sufficiently flexible.
  - But not so when the quantile level is fixed $\rightarrow$ generalized AL distribution.

Two modeling scenarios

- We envision two modeling scenarios (with fixed K throughout):
  - fixed p_k: for settings where one expects specific parts of the response distribution (e.g., the right tail) to be informed by the covariates;
  - the fixed p_k model can also be used when such information is not available, based on a set of equally spaced p_k spanning the unit interval;
  - random p_k: specify only the total number of quantiles, and let the data inform the full configuration of quantile components (both p_k and μ_{p_k}).
- Mixture components need to be flexible distributions parameterized in terms of quantiles.
  - For the random p_k model, the AL distribution is sufficiently flexible.
  - But not so when the quantile level is fixed $\rightarrow$ generalized AL distribution.

Two modeling scenarios

- We envision two modeling scenarios (with fixed K throughout):
  - fixed p_k: for settings where one expects specific parts of the response distribution (e.g., the right tail) to be informed by the covariates;
  - the fixed p_k model can also be used when such information is not available, based on a set of equally spaced p_k spanning the unit interval;
  - random p_k: specify only the total number of quantiles, and let the data inform the full configuration of quantile components (both p_k and μ_{p_k}).
- Mixture components need to be flexible distributions parameterized in terms of quantiles.
  - For the random p_k model, the AL distribution is sufficiently flexible.
  - But not so when the quantile level is fixed $\rightarrow$ generalized AL distribution.

Two modeling scenarios

- We envision two modeling scenarios (with fixed K throughout):
  - fixed p_k: for settings where one expects specific parts of the response distribution (e.g., the right tail) to be informed by the covariates;
  - the fixed p_k model can also be used when such information is not available, based on a set of equally spaced p_k spanning the unit interval;
  - random p_k: specify only the total number of quantiles, and let the data inform the full configuration of quantile components (both p_k and μ_{p_k}).
- Mixture components need to be flexible distributions parameterized in terms of quantiles.
  - For the random p_k model, the AL distribution is sufficiently flexible.
  - But not so when the quantile level is fixed $\rightarrow$ generalized AL distribution.

Asymmetric Laplace distribution

- Asymmetric Laplace (AL) density

$$f_p^{\mathrm{AL}}(y \mid \mu, \sigma) = \frac{p(1-p)}{\sigma} \exp\left\{-\frac{1}{\sigma}\rho_p\left(y-\mu\right)\right\}, \quad y \in \mathbb{R}$$

where $\rho_p(u) = up - u1_{(u<0)}$, $\sigma > 0$ is a scale parameter, $p \in (0,1)$, and $\mu \in \mathbb{R}$ corresponds to the p-th percentile, $\int_{-\infty}^{\mu} f_p^{\mathrm{AL}}(y \mid \mu, \sigma)\mathrm{d}y = p$.

- For $\mu = \boldsymbol{x}'\boldsymbol{\beta}$, maximizing the likelihood w.r.t. $\boldsymbol{\beta}$ under an AL response distribution corresponds to minimizing for $\boldsymbol{\beta}$ the check loss function.
- This property, along with an effective mixture representation, render the AL a popular choice in quantile regression modeling.
- But, the fixed-p AL distribution is very restrictive:
  - the skewness of the error distribution is fully determined by fixing p;
  - the mode of the error density is at 0 for any p.

Asymmetric Laplace distribution

- Asymmetric Laplace (AL) density

$$f_p^{\mathrm{AL}}(y \mid \mu, \sigma) = \frac{p(1-p)}{\sigma} \exp\left\{-\frac{1}{\sigma}\rho_p\left(y-\mu\right)\right\}, \quad y \in \mathbb{R}$$

  where $\rho_p(u) = up - u1_{(u<0)}$, $\sigma > 0$ is a scale parameter, $p \in (0,1)$, and $\mu \in \mathbb{R}$ corresponds to the p-th percentile, $\int_{-\infty}^{\mu} f_p^{\mathrm{AL}}(y \mid \mu, \sigma)\mathrm{d}y = p$.

- For $\mu = \boldsymbol{x}'\boldsymbol{\beta}$, maximizing the likelihood w.r.t. $\boldsymbol{\beta}$ under an AL response distribution corresponds to minimizing for $\boldsymbol{\beta}$ the check loss function.

- This property, along with an effective mixture representation, render the AL a popular choice in quantile regression modeling.

- But, the fixed-p AL distribution is very restrictive:
  - the skewness of the error distribution is fully determined by fixing p;
  - the mode of the error density is at 0 for any p.

Asymmetric Laplace distribution

- Asymmetric Laplace (AL) density

$$f_p^{\mathrm{AL}}(y \mid \mu, \sigma) = \frac{p(1-p)}{\sigma} \exp\left\{-\frac{1}{\sigma}\rho_p\left(y-\mu\right)\right\}, \quad y \in \mathbb{R}$$

  where $\rho_p(u) = up - u1_{(u<0)}$, $\sigma > 0$ is a scale parameter, $p \in (0,1)$, and $\mu \in \mathbb{R}$ corresponds to the p-th percentile, $\int_{-\infty}^{\mu} f_p^{\mathrm{AL}}(y \mid \mu, \sigma)\mathrm{d}y = p$.

- For $\mu = \boldsymbol{x}'\boldsymbol{\beta}$, maximizing the likelihood w.r.t. $\boldsymbol{\beta}$ under an AL response distribution corresponds to minimizing for $\boldsymbol{\beta}$ the check loss function.

- This property, along with an effective mixture representation, render the AL a popular choice in quantile regression modeling.

- But, the fixed-p AL distribution is very restrictive:
  - the skewness of the error distribution is fully determined by fixing p;
  - the mode of the error density is at 0 for any p.

Asymmetric Laplace distribution

- Asymmetric Laplace (AL) density

$$f_p^{\mathrm{AL}}(y \mid \mu, \sigma) = \frac{p(1-p)}{\sigma} \exp\left\{-\frac{1}{\sigma}\rho_p\left(y-\mu\right)\right\}, \quad y \in \mathbb{R}$$

  where $\rho_p(u) = up - u1_{(u<0)}$, $\sigma > 0$ is a scale parameter, $p \in (0,1)$, and $\mu \in \mathbb{R}$ corresponds to the p-th percentile, $\int_{-\infty}^{\mu} f_p^{\mathrm{AL}}(y \mid \mu, \sigma)\mathrm{d}y = p$.

- For $\mu = \boldsymbol{x}'\boldsymbol{\beta}$, maximizing the likelihood w.r.t. $\boldsymbol{\beta}$ under an AL response distribution corresponds to minimizing for $\boldsymbol{\beta}$ the check loss function.

- This property, along with an effective mixture representation, render the AL a popular choice in quantile regression modeling.

- But, the fixed-p AL distribution is very restrictive:
  - the skewness of the error distribution is fully determined by fixing p;
  - the mode of the error density is at 0 for any p.

Generalized AL distribution

- Construction motivated by the AL mixture representation:

$$f_p^{\mathrm{AL}}(y \mid \mu, \sigma) = \int_{\mathbb{R}^+} \mathrm{N}(y \mid \mu + \sigma A(p) z, \sigma^2 B(p) z) \, \mathrm{Exp}(z \mid 1) \, \mathrm{d}z$$

where $A(p) = (1-2p)/\{p(1-p)\}$ and $B(p) = 2/\{p(1-p)\}$.

- Replace the normal kernel with a skew normal density:

$$\frac{2}{\omega} \phi\left(\frac{y-\xi}{\omega}\right) \Phi\left(\frac{\alpha(y-\xi)}{\omega}\right) = \int_{\mathbb{R}^+} \mathrm{N}(y \mid \xi + \tau\alpha s, \tau^2) \mathrm{N}^+(s \mid 0, 1) \, \mathrm{d}s$$

where $\alpha \in \mathbb{R}$ is the skewness parameter, $\tau = \omega(1+\alpha^2)^{-1/2}$, and $\mathrm{N}^+(0,1)$ denotes the standard normal distribution truncated over $\mathbb{R}^+$.

- Generalized asymmetric Laplace (GAL) density:

$$\iint_{\mathbb{R}^+ \times \mathbb{R}^+} \mathrm{N}(y \mid \mu + \sigma\alpha s + \sigma A(p) z, \sigma^2 B(p) z) \, \mathrm{Exp}(z \mid 1) \, \mathrm{N}^+(s \mid 0, 1) \, \mathrm{d}z \, \mathrm{d}s$$

Generalized AL distribution

- Construction motivated by the AL mixture representation:

$$f_p^{\mathrm{AL}}(y \mid \mu, \sigma) = \int_{\mathbb{R}^+} \mathrm{N}(y \mid \mu + \sigma A(p) z, \sigma^2 B(p) z) \, \mathrm{Exp}(z \mid 1) \, \mathrm{d}z$$

where $A(p) = (1-2p)/\{p(1-p)\}$ and $B(p) = 2/\{p(1-p)\}$.

- Replace the normal kernel with a skew normal density:

$$\frac{2}{\omega} \phi\left(\frac{y-\xi}{\omega}\right) \Phi\left(\frac{\alpha(y-\xi)}{\omega}\right) = \int_{\mathbb{R}^+} \mathrm{N}(y \mid \xi + \tau\alpha s, \tau^2) \mathrm{N}^+(s \mid 0, 1) \, \mathrm{d}s$$

where $\alpha \in \mathbb{R}$ is the skewness parameter, $\tau = \omega(1+\alpha^2)^{-1/2}$, and $\mathrm{N}^+(0,1)$ denotes the standard normal distribution truncated over $\mathbb{R}^+$.

- Generalized asymmetric Laplace (GAL) density:

$$\iint_{\mathbb{R}^+ \times \mathbb{R}^+} \mathrm{N}(y \mid \mu + \sigma\alpha s + \sigma A(p) z, \sigma^2 B(p) z) \, \mathrm{Exp}(z \mid 1) \, \mathrm{N}^+(s \mid 0, 1) \, \mathrm{d}z \, \mathrm{d}s$$

Generalized AL distribution

- Construction motivated by the AL mixture representation:

$$f_p^{\mathrm{AL}}(y \mid \mu, \sigma) = \int_{\mathbb{R}^+} \mathrm{N}(y \mid \mu + \sigma A(p) z, \sigma^2 B(p) z)\, \mathrm{Exp}(z \mid 1)\, \mathrm{d}z$$

where $A(p) = (1-2p)/\{p(1-p)\}$ and $B(p) = 2/\{p(1-p)\}$.

- Replace the normal kernel with a skew normal density:

$$\frac{2}{\omega}\, \phi\left(\frac{y-\xi}{\omega}\right) \Phi\left(\frac{\alpha(y-\xi)}{\omega}\right) = \int_{\mathbb{R}^+} \mathrm{N}(y \mid \xi + \tau\alpha s, \tau^2) \mathrm{N}^+(s \mid 0, 1)\, \mathrm{d}s$$

where $\alpha \in \mathbb{R}$ is the skewness parameter, $\tau = \omega(1+\alpha^2)^{-1/2}$, and $\mathrm{N}^+(0,1)$ denotes the standard normal distribution truncated over $\mathbb{R}^+$.

- Generalized asymmetric Laplace (GAL) density:

$$\iint_{\mathbb{R}^+ \times \mathbb{R}^+} \mathrm{N}(y \mid \mu + \sigma\alpha s + \sigma A(p) z, \sigma^2 B(p) z)\, \mathrm{Exp}(z \mid 1)\, \mathrm{N}^+(s \mid 0, 1)\, \mathrm{d}z\, \mathrm{d}s$$

Generalized AL distribution

- GAL density $f(y \mid p, \mu, \sigma, \alpha) =$

$$\iint_{\mathbb{R}^+ \times \mathbb{R}^+} \mathrm{N}(y \mid \mu + \sigma\alpha s + \sigma A(p) z, \sigma^2 B(p) z)\, \mathrm{Exp}(z \mid 1)\, \mathrm{N}^+(s \mid 0, 1)\, \mathrm{d}z\, \mathrm{d}s$$

  - When $\alpha = 0$, the GAL density reduces to the AL density.
  - Integrating over $z \rightarrow$ generalized inverse-Gaussian density $\rightarrow$ integrating over $s \rightarrow$ (complex) closed-form expression for $f(y \mid p, \mu, \sigma, \alpha)$.
  - Hierarchical mixture representation of the GAL density suffices for study of model properties and for model fitting.

- p_0-th quantile, for $p_0 \in (0, 1)$? Setting $\gamma = \{I(\alpha > 0) - p\}|\alpha|$,

$$\int_{-\infty}^{\mu} f(y \mid p, \mu, \sigma, \gamma)\, \mathrm{d}y = p\, g(\gamma) \quad \text{with} \quad g(\gamma) = 2\Phi(-|\gamma|) \exp(\gamma^2/2).$$

  - $g(\gamma)$ is increasing in $\mathbb{R}^-$, and decreasing in $\mathbb{R}^+$;
  - for any γ, unique solution for p such that $\int_{-\infty}^{\mu} f(y \mid p, \mu, \sigma, \gamma)\, \mathrm{d}y = p_0$.

Generalized AL distribution

- GAL density $f(y \mid p, \mu, \sigma, \alpha) =$

$$\iint_{\mathbb{R}^+ \times \mathbb{R}^+} \mathrm{N}(y \mid \mu + \sigma\alpha s + \sigma A(p) z, \sigma^2 B(p) z)\, \mathrm{Exp}(z \mid 1)\, \mathrm{N}^+(s \mid 0, 1)\, \mathrm{d}z\, \mathrm{d}s$$

  - When $\alpha = 0$, the GAL density reduces to the AL density.
  - Integrating over $z \rightarrow$ generalized inverse-Gaussian density $\rightarrow$ integrating over $s \rightarrow$ (complex) closed-form expression for $f(y \mid p, \mu, \sigma, \alpha)$.
  - Hierarchical mixture representation of the GAL density suffices for study of model properties and for model fitting.

- p_0-th quantile, for $p_0 \in (0, 1)$? Setting $\gamma = \{I(\alpha > 0) - p\}|\alpha|$,

$$\int_{-\infty}^{\mu} f(y \mid p, \mu, \sigma, \gamma)\, \mathrm{d}y = p\, g(\gamma) \quad \text{with} \quad g(\gamma) = 2\Phi(-|\gamma|) \exp(\gamma^2/2).$$

  - $g(\gamma)$ is increasing in $\mathbb{R}^-$, and decreasing in $\mathbb{R}^+$;
  - for any γ, unique solution for p such that $\int_{-\infty}^{\mu} f(y \mid p, \mu, \sigma, \gamma)\, \mathrm{d}y = p_0$.

Generalized AL distribution

- Reparameterize (p, α) to $(p_0, \gamma) \longrightarrow$ for any fixed $p_0 \in (0, 1)$, the p_0-th quantile of $f(y \mid p_0, \mu, \sigma, \gamma) \equiv f_{p_0}(y \mid \mu, \sigma, \gamma)$ is μ.

- Quantile-fixed (fixed p_0) GAL density $f_{p_0}(y \mid \mu, \sigma, \gamma) =$

$$\iint_{\mathbb{R}^+ \times \mathbb{R}^+} \mathrm{N}(y \mid \mu + \sigma C|\gamma|s + \sigma A z, \sigma^2 B z)\, \mathrm{Exp}(z \mid 1)\, \mathrm{N}^+(s \mid 0, 1)\, \mathrm{d}z\, \mathrm{d}s$$

  - A, B and C are all functions of γ (and p_0).
  - Y has density $f_{p_0}(\mu, \sigma, \gamma)$ if-f $(Y - \mu)/\sigma$ has density $f_{p_0}(0, 1, \gamma) \rightarrow \mu$ is a location parameter (the p_0-th quantile) and σ is a scale parameter.
  - $\gamma \in (L, U)$, where L is the negative root of $g(\gamma) = 1 - p_0$ and U is the positive root of $g(\gamma) = p_0$.
  - γ is a shape parameter that allows for more flexible densities than the AL.

Generalized AL distribution

- Reparameterize (p, α) to $(p_0, \gamma) \longrightarrow$ for any fixed $p_0 \in (0, 1)$, the p_0-th quantile of $f(y \mid p_0, \mu, \sigma, \gamma) \equiv f_{p_0}(y \mid \mu, \sigma, \gamma)$ is μ.

- Quantile-fixed (fixed p_0) GAL density $f_{p_0}(y \mid \mu, \sigma, \gamma) =$

$$\iint_{\mathbb{R}^+ \times \mathbb{R}^+} \mathrm{N}(y \mid \mu + \sigma C|\gamma|s + \sigma A z, \sigma^2 B z)\, \mathrm{Exp}(z \mid 1)\, \mathrm{N}^+(s \mid 0, 1)\, \mathrm{d}z\, \mathrm{d}s$$

  - A, B and C are all functions of γ (and p_0).
  - Y has density $f_{p_0}(\mu, \sigma, \gamma)$ if-f $(Y - \mu)/\sigma$ has density $f_{p_0}(0, 1, \gamma) \rightarrow \mu$ is a location parameter (the p_0-th quantile) and σ is a scale parameter.
  - $\gamma \in (L, U)$, where L is the negative root of $g(\gamma) = 1 - p_0$ and U is the positive root of $g(\gamma) = p_0$.
  - γ is a shape parameter that allows for more flexible densities than the AL.

Generalized AL distribution

- Reparameterize (p, α) to (p_0, γ) $\longrightarrow$ for any fixed $p_0 \in (0, 1)$, the p_0-th quantile of $f(y \mid p_0, \mu, \sigma, \gamma) \equiv f_{p_0}(y \mid \mu, \sigma, \gamma)$ is μ.

- Quantile-fixed (fixed p_0) GAL density $f_{p_0}(y \mid \mu, \sigma, \gamma) =$

$$\iint_{\mathbb{R}^+ \times \mathbb{R}^+} \mathrm{N}(y \mid \mu + \sigma C|\gamma|s + \sigma A z, \sigma^2 B z)\, \mathrm{Exp}(z \mid 1)\, \mathrm{N}^+(s \mid 0, 1)\, \mathrm{d}z\, \mathrm{d}s$$

  - A, B and C are all functions of γ (and p_0).
  - Y has density $f_{p_0}(\mu, \sigma, \gamma)$ if-f $(Y - \mu)/\sigma$ has density $f_{p_0}(0, 1, \gamma)$ $\rightarrow$ μ is a location parameter (the p_0-th quantile) and σ is a scale parameter.
  - $\gamma \in (L, U)$, where L is the negative root of $g(\gamma) = 1 - p_0$ and U is the positive root of $g(\gamma) = p_0$.
  - γ is a shape parameter that allows for more flexible densities than the AL.

Generalized AL distribution

- Reparameterize (p, α) to (p_0, γ) $\longrightarrow$ for any fixed $p_0 \in (0, 1)$, the p_0-th quantile of $f(y \mid p_0, \mu, \sigma, \gamma) \equiv f_{p_0}(y \mid \mu, \sigma, \gamma)$ is μ.

- Quantile-fixed (fixed p_0) GAL density $f_{p_0}(y \mid \mu, \sigma, \gamma) =$

$$\iint_{\mathbb{R}^+ \times \mathbb{R}^+} \mathrm{N}(y \mid \mu + \sigma C|\gamma|s + \sigma A z, \sigma^2 B z)\, \mathrm{Exp}(z \mid 1)\, \mathrm{N}^+(s \mid 0, 1)\, \mathrm{d}z\, \mathrm{d}s$$

  - A, B and C are all functions of γ (and p_0).
  - Y has density $f_{p_0}(\mu, \sigma, \gamma)$ if-f $(Y - \mu)/\sigma$ has density $f_{p_0}(0, 1, \gamma)$ $\rightarrow$ μ is a location parameter (the p_0-th quantile) and σ is a scale parameter.
  - $\gamma \in (L, U)$, where L is the negative root of $g(\gamma) = 1 - p_0$ and U is the positive root of $g(\gamma) = p_0$.
  - γ is a shape parameter that allows for more flexible densities than the AL.

Generalized AL distribution

- Reparameterize (p, α) to (p_0, γ) $\longrightarrow$ for any fixed $p_0 \in (0, 1)$, the p_0-th quantile of $f(y \mid p_0, \mu, \sigma, \gamma) \equiv f_{p_0}(y \mid \mu, \sigma, \gamma)$ is μ.

- Quantile-fixed (fixed p_0) GAL density $f_{p_0}(y \mid \mu, \sigma, \gamma) =$

$$\iint_{\mathbb{R}^+ \times \mathbb{R}^+} \mathrm{N}(y \mid \mu + \sigma C|\gamma|s + \sigma A z, \sigma^2 B z)\, \mathrm{Exp}(z \mid 1)\, \mathrm{N}^+(s \mid 0, 1)\, \mathrm{d}z\, \mathrm{d}s$$

  - A, B and C are all functions of γ (and p_0).
  - Y has density $f_{p_0}(\mu, \sigma, \gamma)$ if-f $(Y - \mu)/\sigma$ has density $f_{p_0}(0, 1, \gamma)$ $\rightarrow$ μ is a location parameter (the p_0-th quantile) and σ is a scale parameter.
  - $\gamma \in (L, U)$, where L is the negative root of $g(\gamma) = 1 - p_0$ and U is the positive root of $g(\gamma) = p_0$.
  - γ is a shape parameter that allows for more flexible densities than the AL.

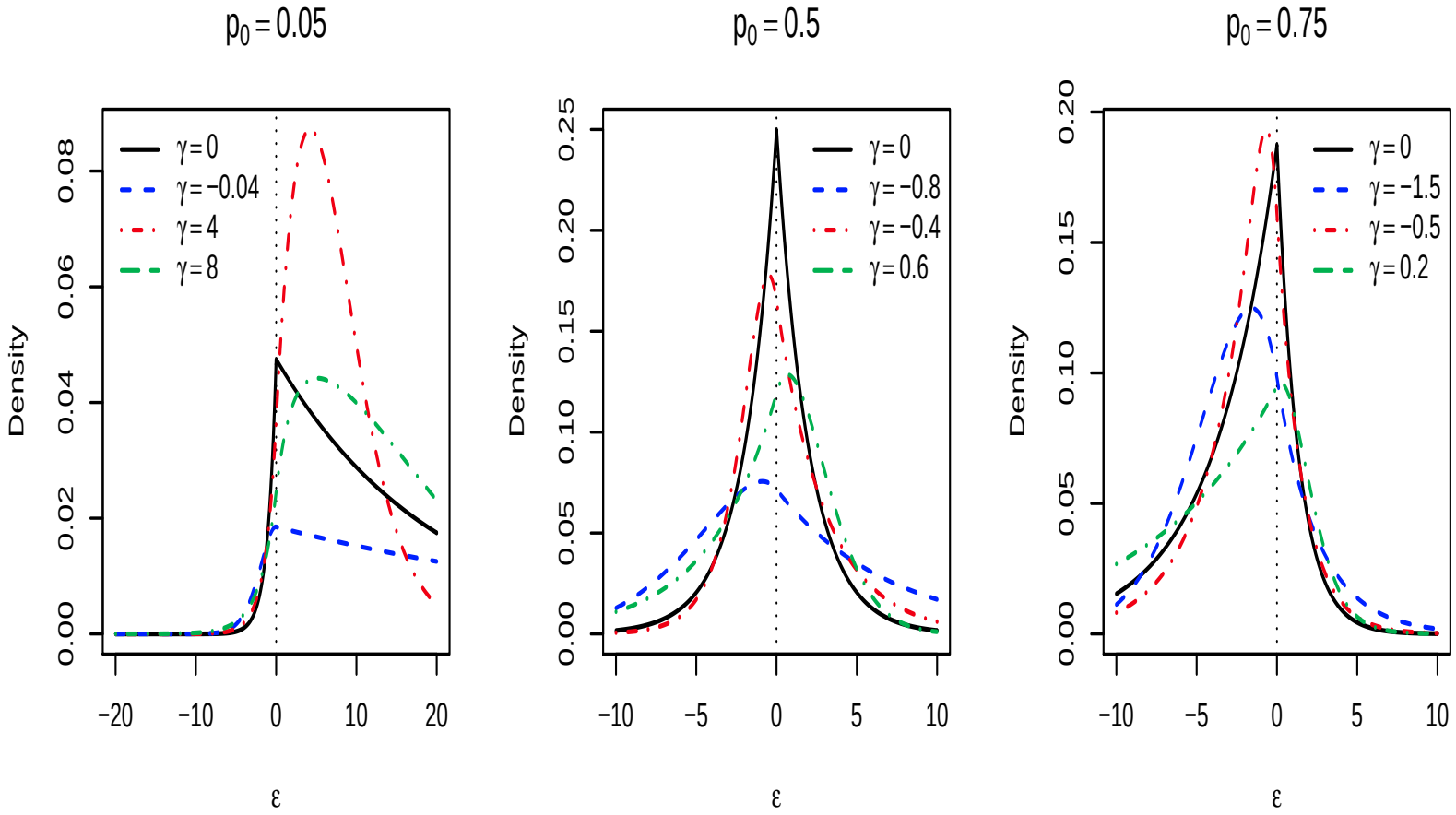


Figure: Quantile level fixed GAL($\mu = 0, \sigma = 1, \gamma$) densities, with $p_0 = 0.05$ ($\gamma \in (-0.07, 15.90)$), $p_0 = 0.5$ ($\gamma \in (-1.09, 1.09)$), and $p_0 = 0.75$ ($\gamma \in (-2.90, 0.39)$).

Quantile regression with regularization

- The extension of the AL to the GAL distribution from the check loss function perspective.

- Marginalize over the $z_i \longrightarrow \pi(\boldsymbol{\beta}, \gamma, \sigma, s_1, ..., s_n \mid \text{data}) \longrightarrow$ posterior full conditional for $\boldsymbol{\beta}$:

$$\pi(\boldsymbol{\beta} \mid \gamma, \sigma, s_1, ..., s_n, \text{data}) \propto \pi(\boldsymbol{\beta}) \exp\left\{-\frac{1}{\sigma}\sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} - \sigma H(\gamma) s_i)\right\}$$

  - $H(\gamma) = \gamma g(\gamma)/\{g(\gamma) - |p_0 - I(\gamma < 0)|\}$
  - $p = I(\gamma < 0) + \{[p_0 - I(\gamma < 0)]/g(\gamma)\}$
  - p_0 the probability associated with the quantile modeled through $\boldsymbol{x}_i^T\boldsymbol{\beta}$.
  - For $\gamma = 0$ (AL errors) $\longrightarrow$ check loss function with $p = p_0$.

Quantile regression with regularization

- The extension of the AL to the GAL distribution from the check loss function perspective.

- Marginalize over the $z_i \longrightarrow \pi(\boldsymbol{\beta}, \gamma, \sigma, s_1, ..., s_n \mid \text{data}) \longrightarrow$ posterior full conditional for $\boldsymbol{\beta}$:

$$\pi(\boldsymbol{\beta} \mid \gamma, \sigma, s_1, ..., s_n, \text{data}) \propto \pi(\boldsymbol{\beta}) \exp\left\{-\frac{1}{\sigma}\sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} - \sigma H(\gamma) s_i)\right\}$$

  - $H(\gamma) = \gamma g(\gamma)/\{g(\gamma) - |p_0 - I(\gamma < 0)|\}$
  - $p = I(\gamma < 0) + \{[p_0 - I(\gamma < 0)]/g(\gamma)\}$
  - p_0 the probability associated with the quantile modeled through $\boldsymbol{x}_i^T\boldsymbol{\beta}$.
  - For $\gamma = 0$ (AL errors) $\longrightarrow$ check loss function with $p = p_0$.

Quantile regression with regularization

- The extension of the AL to the GAL distribution from the check loss function perspective.

- Marginalize over the $z_i \longrightarrow \pi(\boldsymbol{\beta}, \gamma, \sigma, s_1, ..., s_n \mid \text{data}) \longrightarrow$ posterior full conditional for $\boldsymbol{\beta}$:

$$\pi(\boldsymbol{\beta} \mid \gamma, \sigma, s_1, ..., s_n, \text{data}) \propto \pi(\boldsymbol{\beta}) \exp\left\{-\frac{1}{\sigma}\sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} - \sigma H(\gamma) s_i)\right\}$$

  - $H(\gamma) = \gamma g(\gamma)/\{g(\gamma) - |p_0 - I(\gamma < 0)|\}$
  - $p = I(\gamma < 0) + \{[p_0 - I(\gamma < 0)]/g(\gamma)\}$
  - p_0 the probability associated with the quantile modeled through $\boldsymbol{x}_i^T\boldsymbol{\beta}$.
  - For $\gamma = 0$ (AL errors) $\longrightarrow$ check loss function with $p = p_0$.

Quantile regression with regularization

- Adjusted loss function $\sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta} - \sigma H(\gamma) s_i)$
  - Positive-valued latent variables s_i can be viewed as response-specific weights that are adjusted by real-valued coefficient $H(\gamma)$, which is fully specified through the shape parameter γ.
  - Real-valued, response-specific terms $\sigma H(\gamma) s_i$ reflect on the estimation of $\boldsymbol{\beta}$ the effect of outlying observations relative to the AL distribution.

- Different versions of regularized quantile regression under different priors for $\boldsymbol{\beta}$, working with AL errors (Li et al., 2010).
  - Lasso regularized quantile regression $\longrightarrow$ hierarchical Laplace prior, $\pi(\boldsymbol{\beta} \mid \sigma, \lambda) = \prod_j 0.5\, \lambda\, \sigma^{-1} \exp(-\lambda \sigma^{-1} |\beta_j|)$

- A broader framework for exploring regularization by adjusting the loss function (through the response distribution) in addition to the penalty term (through the prior for the regression coefficients).

Quantile regression with regularization

- Adjusted loss function $\sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta} - \sigma H(\gamma) s_i)$
  - Positive-valued latent variables s_i can be viewed as response-specific weights that are adjusted by real-valued coefficient $H(\gamma)$, which is fully specified through the shape parameter γ.
  - Real-valued, response-specific terms $\sigma H(\gamma) s_i$ reflect on the estimation of $\boldsymbol{\beta}$ the effect of outlying observations relative to the AL distribution.

- Different versions of regularized quantile regression under different priors for $\boldsymbol{\beta}$, working with AL errors (Li et al., 2010).
  - Lasso regularized quantile regression $\longrightarrow$ hierarchical Laplace prior, $\pi(\boldsymbol{\beta} \mid \sigma, \lambda) = \prod_j 0.5\, \lambda\, \sigma^{-1} \exp(-\lambda \sigma^{-1} |\beta_j|)$

- A broader framework for exploring regularization by adjusting the loss function (through the response distribution) in addition to the penalty term (through the prior for the regression coefficients).

Quantile regression with regularization

- Adjusted loss function $\sum_{i=1}^{n} \rho_p(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta} - \sigma H(\gamma) s_i)$
  - Positive-valued latent variables s_i can be viewed as response-specific weights that are adjusted by real-valued coefficient $H(\gamma)$, which is fully specified through the shape parameter γ.
  - Real-valued, response-specific terms $\sigma H(\gamma) s_i$ reflect on the estimation of $\boldsymbol{\beta}$ the effect of outlying observations relative to the AL distribution.

- Different versions of regularized quantile regression under different priors for $\boldsymbol{\beta}$, working with AL errors (Li et al., 2010).
  - Lasso regularized quantile regression $\longrightarrow$ hierarchical Laplace prior, $\pi(\boldsymbol{\beta} \mid \sigma, \lambda) = \prod_j 0.5\, \lambda\, \sigma^{-1} \exp(-\lambda \sigma^{-1} |\beta_j|)$

- A broader framework for exploring regularization by adjusting the loss function (through the response distribution) in addition to the penalty term (through the prior for the regression coefficients).

GAL mixture model (fixed p_k)

- Use GAL densities for the mixture components:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \sigma, \gamma_{p_k})$$

- Specify the values for $0 < p_1 < \ldots < p_K < 1$.
- Conditional truncated normal priors for $\mu_{p_1} < \ldots < \mu_{p_K}$.
- Rescaled Beta priors for the γ_{p_k} (*default* uniform).
- Hierarchical Laplace prior for $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta} \mid \sigma, \lambda) = \prod_{j=1}^{d} \frac{\lambda}{2\sigma} \exp(-\frac{\lambda}{\sigma}|\beta_j|)$.
- Mixture weights defined through increments of a c.d.f. G (on $(0, p_K)$):

$$\omega_1 = G(p_1), \;\; \omega_k = G(p_k) - G(p_{k-1}), \; k = 2, \ldots, K$$

  where G is assigned a Dirichlet process prior.

- MCMC: one set of latent variables for the mixture, two sets for the GAL densities $\rightarrow$ M-H steps for the γ_{p_k}; Gibbs sampling for all other parameters.

GAL mixture model (fixed p_k)

- Use GAL densities for the mixture components:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \sigma, \gamma_{p_k})$$

  - Specify the values for $0 < p_1 < \ldots < p_K < 1$.
  - Conditional truncated normal priors for $\mu_{p_1} < \ldots < \mu_{p_K}$.
  - Rescaled Beta priors for the γ_{p_k} (*default* uniform).
  - Hierarchical Laplace prior for $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta} \mid \sigma, \lambda) = \prod_{j=1}^{d} \frac{\lambda}{2\sigma} \exp(-\frac{\lambda}{\sigma}|\beta_j|)$.
  - Mixture weights defined through increments of a c.d.f. G (on $(0, p_K)$):

$$\omega_1 = G(p_1), \;\; \omega_k = G(p_k) - G(p_{k-1}), \; k = 2, \ldots, K$$

    where G is assigned a Dirichlet process prior.

  - MCMC: one set of latent variables for the mixture, two sets for the GAL densities $\rightarrow$ M-H steps for the γ_{p_k}; Gibbs sampling for all other parameters.

GAL mixture model (fixed p_k)

- Use GAL densities for the mixture components:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \sigma, \gamma_{p_k})$$

  - Specify the values for $0 < p_1 < \ldots < p_K < 1$.
  - Conditional truncated normal priors for $\mu_{p_1} < \ldots < \mu_{p_K}$.
  - Rescaled Beta priors for the γ_{p_k} (*default* uniform).
  - Hierarchical Laplace prior for $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta} \mid \sigma, \lambda) = \prod_{j=1}^{d} \frac{\lambda}{2\sigma} \exp(-\frac{\lambda}{\sigma}|\beta_j|)$.
  - Mixture weights defined through increments of a c.d.f. G (on $(0, p_K)$):

$$\omega_1 = G(p_1), \;\; \omega_k = G(p_k) - G(p_{k-1}), \; k = 2, \ldots, K$$

    where G is assigned a Dirichlet process prior.

  - MCMC: one set of latent variables for the mixture, two sets for the GAL densities $\rightarrow$ M-H steps for the γ_{p_k}; Gibbs sampling for all other parameters.

GAL mixture model (fixed p_k)

- Use GAL densities for the mixture components:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \sigma, \gamma_{p_k})$$

  - Specify the values for $0 < p_1 < \ldots < p_K < 1$.
  - Conditional truncated normal priors for $\mu_{p_1} < \ldots < \mu_{p_K}$.
  - Rescaled Beta priors for the γ_{p_k} (*default* uniform).
  - Hierarchical Laplace prior for $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta} \mid \sigma, \lambda) = \prod_{j=1}^{d} \frac{\lambda}{2\sigma} \exp(-\frac{\lambda}{\sigma}|\beta_j|)$.
  - Mixture weights defined through increments of a c.d.f. G (on $(0, p_K)$):

$$\omega_1 = G(p_1), \;\; \omega_k = G(p_k) - G(p_{k-1}), \; k = 2, \ldots, K$$

    where G is assigned a Dirichlet process prior.

  - MCMC: one set of latent variables for the mixture, two sets for the GAL densities $\rightarrow$ M-H steps for the γ_{p_k}; Gibbs sampling for all other parameters.

GAL mixture model (fixed p_k)

- Use GAL densities for the mixture components:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \sigma, \gamma_{p_k})$$

  - Specify the values for $0 < p_1 < \ldots < p_K < 1$.
  - Conditional truncated normal priors for $\mu_{p_1} < \ldots < \mu_{p_K}$.
  - Rescaled Beta priors for the γ_{p_k} (*default* uniform).
  - Hierarchical Laplace prior for $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta} \mid \sigma, \lambda) = \prod_{j=1}^{d} \frac{\lambda}{2\sigma} \exp(-\frac{\lambda}{\sigma}|\beta_j|)$.
  - Mixture weights defined through increments of a c.d.f. G (on $(0, p_K)$):

$$\omega_1 = G(p_1), \;\; \omega_k = G(p_k) - G(p_{k-1}), \; k = 2, \ldots, K$$

    where G is assigned a Dirichlet process prior.

  - MCMC: one set of latent variables for the mixture, two sets for the GAL densities $\rightarrow$ M-H steps for the γ_{p_k}; Gibbs sampling for all other parameters.

GAL mixture model (fixed p_k)

- Use GAL densities for the mixture components:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \sigma, \gamma_{p_k})$$

  - Specify the values for $0 < p_1 < \ldots < p_K < 1$.
  - Conditional truncated normal priors for $\mu_{p_1} < \ldots < \mu_{p_K}$.
  - Rescaled Beta priors for the γ_{p_k} (*default* uniform).
  - Hierarchical Laplace prior for $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta} \mid \sigma, \lambda) = \prod_{j=1}^{d} \frac{\lambda}{2\sigma} \exp(-\frac{\lambda}{\sigma}|\beta_j|)$.
  - Mixture weights defined through increments of a c.d.f. G (on $(0, p_K)$):

$$\omega_1 = G(p_1), \;\; \omega_k = G(p_k) - G(p_{k-1}), \; k = 2, \ldots, K$$

    where G is assigned a Dirichlet process prior.

- MCMC: one set of latent variables for the mixture, two sets for the GAL densities $\rightarrow$ M-H steps for the γ_{p_k}; Gibbs sampling for all other parameters.

GAL mixture model (fixed p_k)

- Use GAL densities for the mixture components:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \sigma, \gamma_{p_k})$$

  - Specify the values for $0 < p_1 < \ldots < p_K < 1$.
  - Conditional truncated normal priors for $\mu_{p_1} < \ldots < \mu_{p_K}$.
  - Rescaled Beta priors for the γ_{p_k} (*default* uniform).
  - Hierarchical Laplace prior for $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta} \mid \sigma, \lambda) = \prod_{j=1}^{d} \frac{\lambda}{2\sigma} \exp(-\frac{\lambda}{\sigma}|\beta_j|)$.
  - Mixture weights defined through increments of a c.d.f. G (on $(0, p_K)$):

$$\omega_1 = G(p_1), \;\; \omega_k = G(p_k) - G(p_{k-1}), \; k = 2, \ldots, K$$

    where G is assigned a Dirichlet process prior.

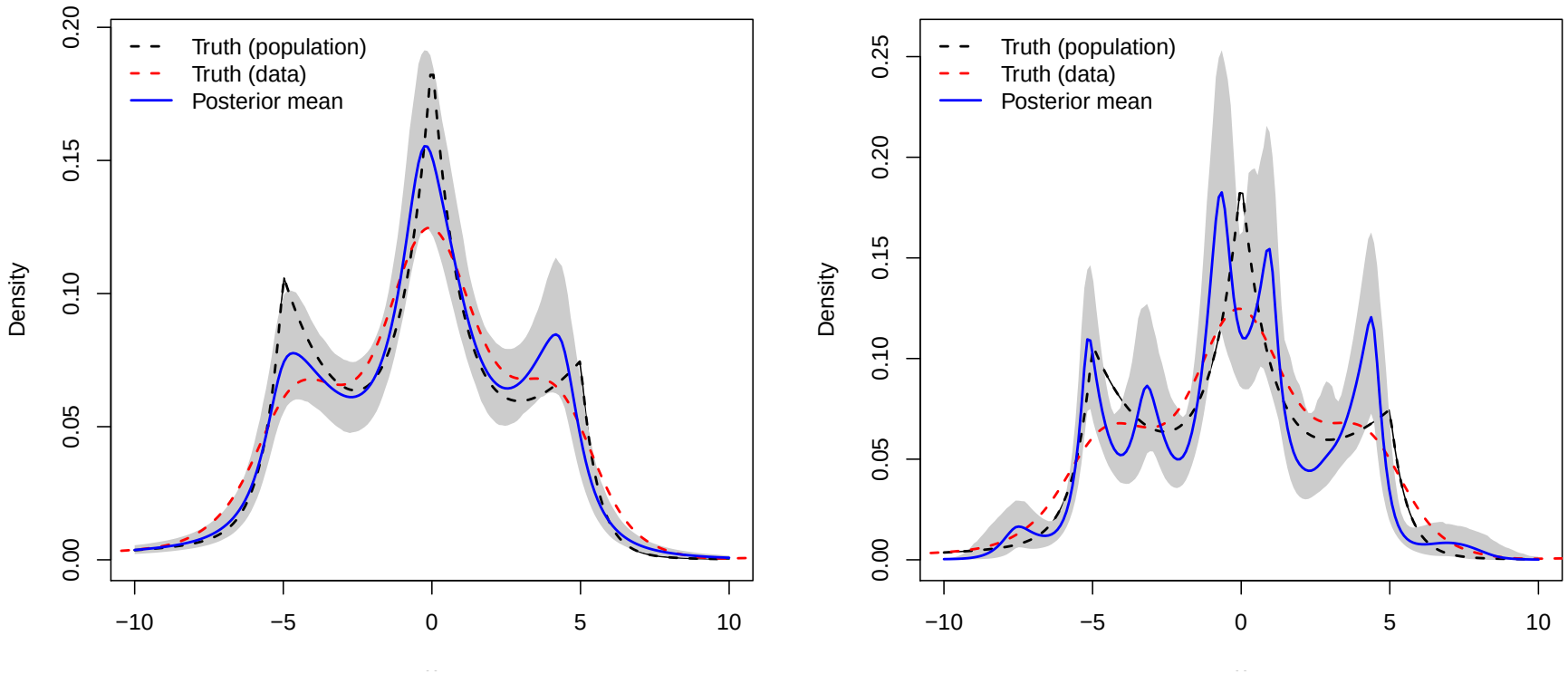
  - MCMC: one set of latent variables for the mixture, two sets for the GAL densities $\rightarrow$ M-H steps for the γ_{p_k}; Gibbs sampling for all other parameters.

Synthetic data examples

- Different data sets simulated from $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$, where:
  - $\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)'$;
  - $\boldsymbol{x}_i$ generated independently from a $\mathrm{N}_8(0, \Sigma)$ distribution, with (i,j)th covariance element $0.5^{|i-j|}$, for $1 \leq i, j \leq 8$.
- Different scenarios for the error distribution:
  - mixture of three AL components (to highlight the benefits of the GAL mixture kernel), with $n = 600$;
  - normal and skew-normal distributions, with $n = 500$ in each case.

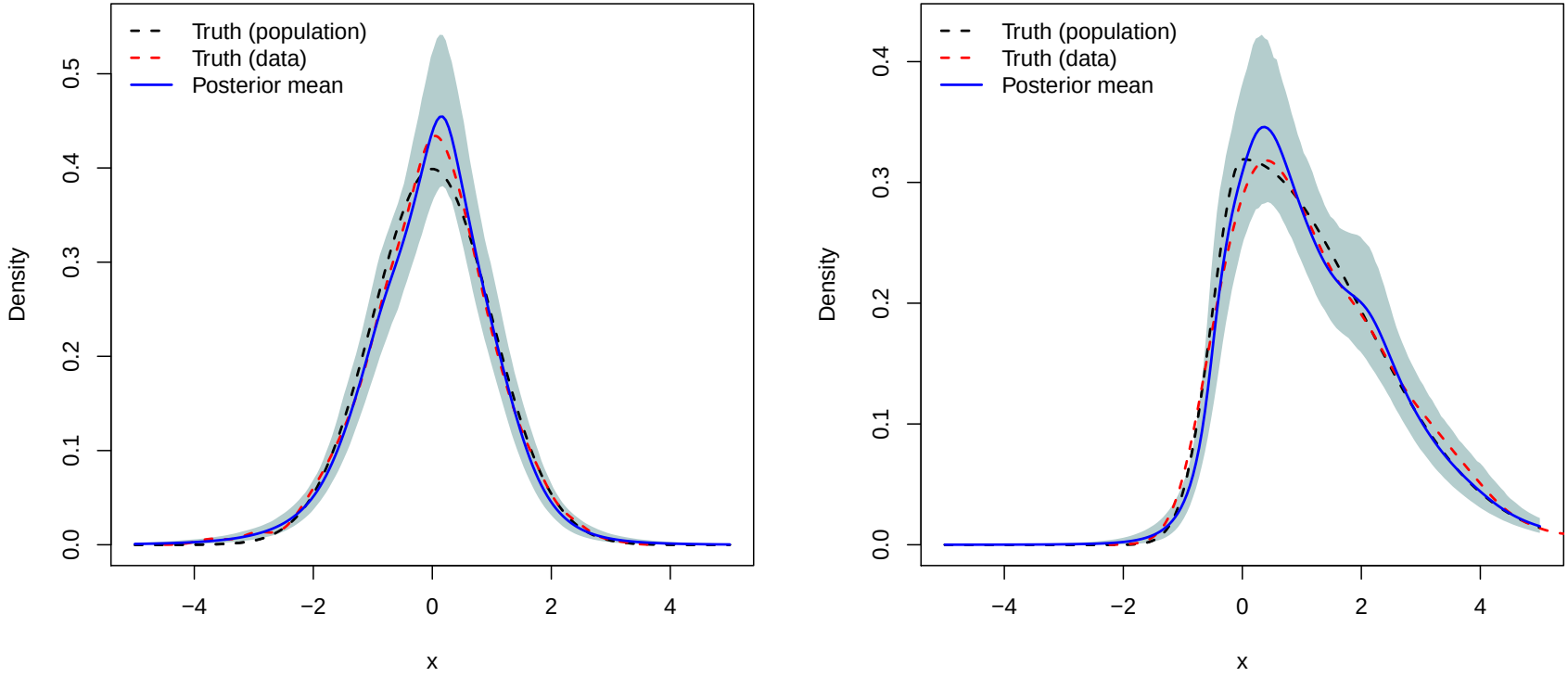
Synthetic data examples

- Different data sets simulated from $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$, where:
  - $\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)'$;
  - $\boldsymbol{x}_i$ generated independently from a $\mathrm{N}_8(0, \Sigma)$ distribution, with (i,j)th covariance element $0.5^{|i-j|}$, for $1 \leq i, j \leq 8$.
- Different scenarios for the error distribution:
  - mixture of three AL components (to highlight the benefits of the GAL mixture kernel), with $n = 600$;
  - normal and skew-normal distributions, with $n = 500$ in each case.



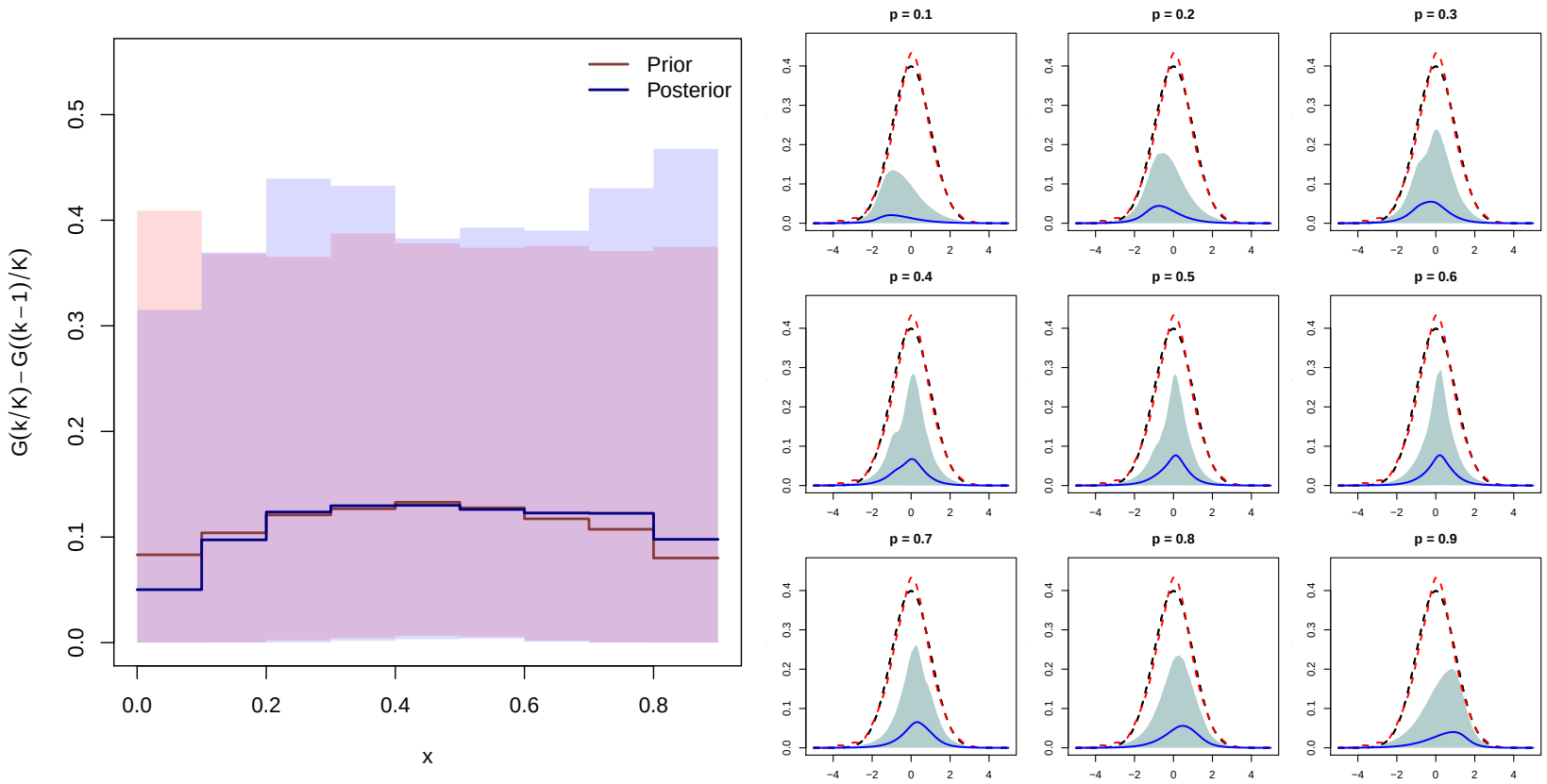
(a) GAL mixture (b) AL mixture

Figure: Synthetic data from a mixture of three AL components. Posterior mean and 95% interval estimates for the error density under the GAL and AL mixture models. In both cases, $p_k = k/10$, for $k = 1, \ldots, 9$.



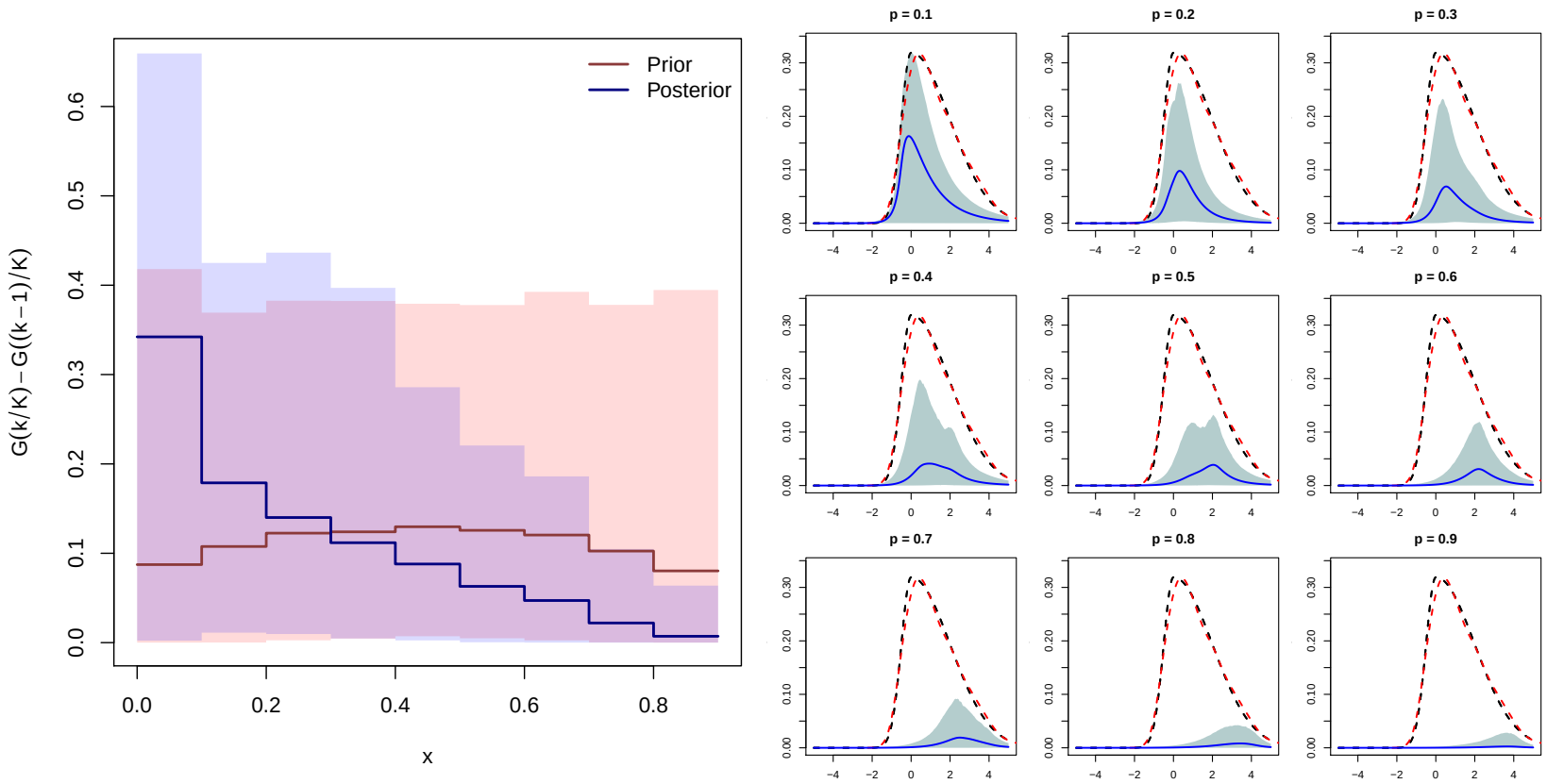
(a) Normal errors (b) Skew-normal errors

Figure: Synthetic data from normal and skew-normal distributions. Posterior mean and 95% interval estimates for the error density under the GAL mixture model (with fixed $p_k = k/10$, for $k = 1, \ldots, 9$).



(a) Mixture weights (b) Weighted mixture components

Figure: Synthetic data from normal distribution. Prior and posterior for the mixture weights, and posterior mean and 95% interval estimates for the error density weighted components, under the GAL mixture model (with fixed $p_k = k/10$, for $k = 1, \ldots, 9$).



(a) Mixture weights (b) Weighted mixture components

Figure: Synthetic data from skew-normal distribution. Prior and posterior for the mixture weights, and posterior mean and 95% interval estimates for the error density weighted components, under the GAL mixture model (with fixed $p_k = k/10$, for $k = 1, \ldots, 9$).

AL mixture model (random p_k)

- Mixture with AL components:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}^{\mathrm{AL}}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \sigma)$$

  - Random $0 < p_1 < \ldots < p_K < 1$ generated from a Poisson process on $(0, 1)$ conditioning on K (uniform subject to the monotonicity restriction).
  - Similar priors with before for the $\mu_{p_1} < \ldots < \mu_{p_K}$, for $\boldsymbol{\beta}$, and for the mixture weights.

- Illustration (and comparison with GAL mixture) using synthetic data from a skew-normal distribution ($n = 200$).

AL mixture model (random p_k)

- Mixture with AL components:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}^{\mathrm{AL}}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \sigma)$$

  - Random $0 < p_1 < \ldots < p_K < 1$ generated from a Poisson process on $(0, 1)$ conditioning on K (uniform subject to the monotonicity restriction).
  - Similar priors with before for the $\mu_{p_1} < \ldots < \mu_{p_K}$, for $\boldsymbol{\beta}$, and for the mixture weights.

- Illustration (and comparison with GAL mixture) using synthetic data from a skew-normal distribution ($n = 200$).

AL mixture model (random p_k)

- Mixture with AL components:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}^{\mathrm{AL}}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \sigma)$$

  - Random $0 < p_1 < \ldots < p_K < 1$ generated from a Poisson process on $(0, 1)$ conditioning on K (uniform subject to the monotonicity restriction).
  - Similar priors with before for the $\mu_{p_1} < \ldots < \mu_{p_K}$, for $\boldsymbol{\beta}$, and for the mixture weights.

- Illustration (and comparison with GAL mixture) using synthetic data from a skew-normal distribution ($n = 200$).

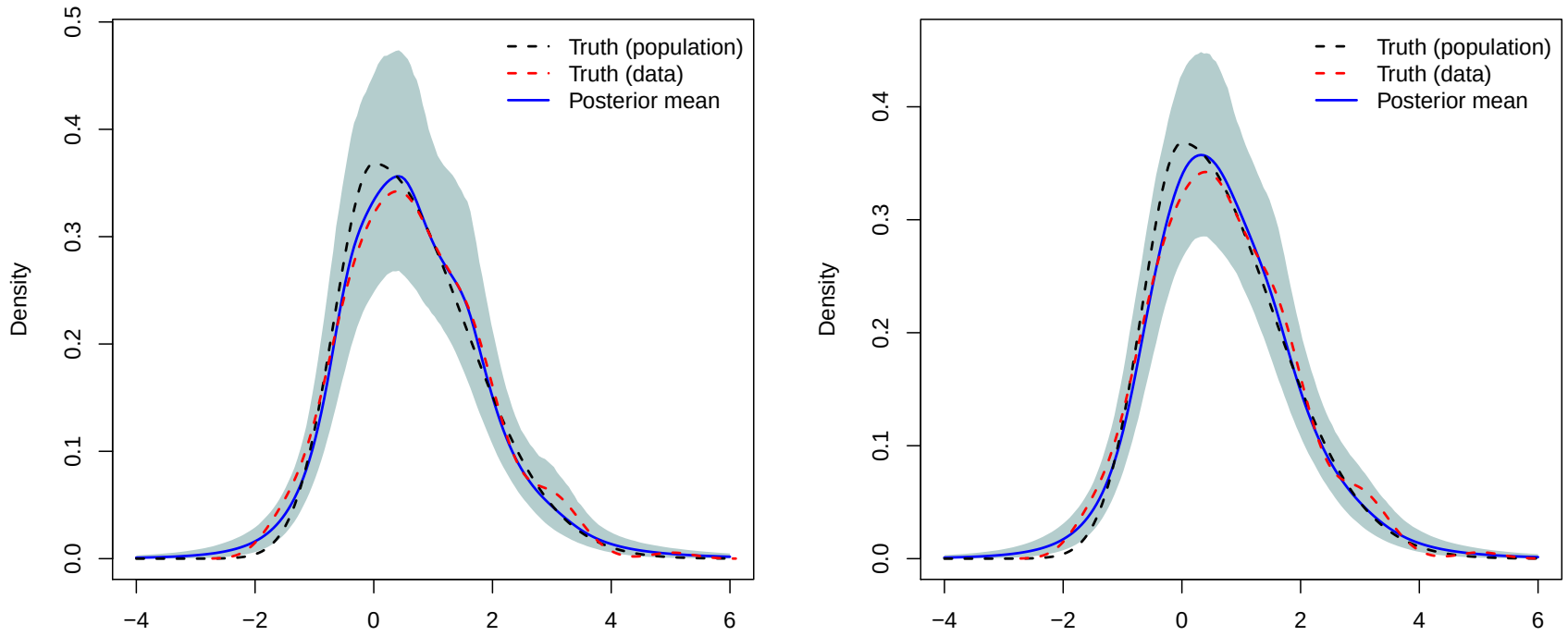
AL mixture model (random p_k)

- Mixture with AL components:

$$f(y \mid \boldsymbol{x}) = \sum_{k=1}^{K} \omega_k \, f_{p_k}^{\mathrm{AL}}(y \mid \mu_{p_k} + \boldsymbol{x}'\boldsymbol{\beta}, \sigma)$$

  - Random $0 < p_1 < \ldots < p_K < 1$ generated from a Poisson process on $(0, 1)$ conditioning on K (uniform subject to the monotonicity restriction).
  - Similar priors with before for the $\mu_{p_1} < \ldots < \mu_{p_K}$, for $\boldsymbol{\beta}$, and for the mixture weights.

- Illustration (and comparison with GAL mixture) using synthetic data from a skew-normal distribution ($n = 200$).



(a) AL mixture, random p_k

(b) GAL mixture, fixed p_k

Figure: Synthetic data from skew-normal distribution. Posterior mean and 95% interval estimates for the error density under the AL mixture with $K = 5$ and random p_k, and the GAL mixture with fixed $p_k = \{0.1, 0.25, 0.5, 0.75, 0.9\}$.

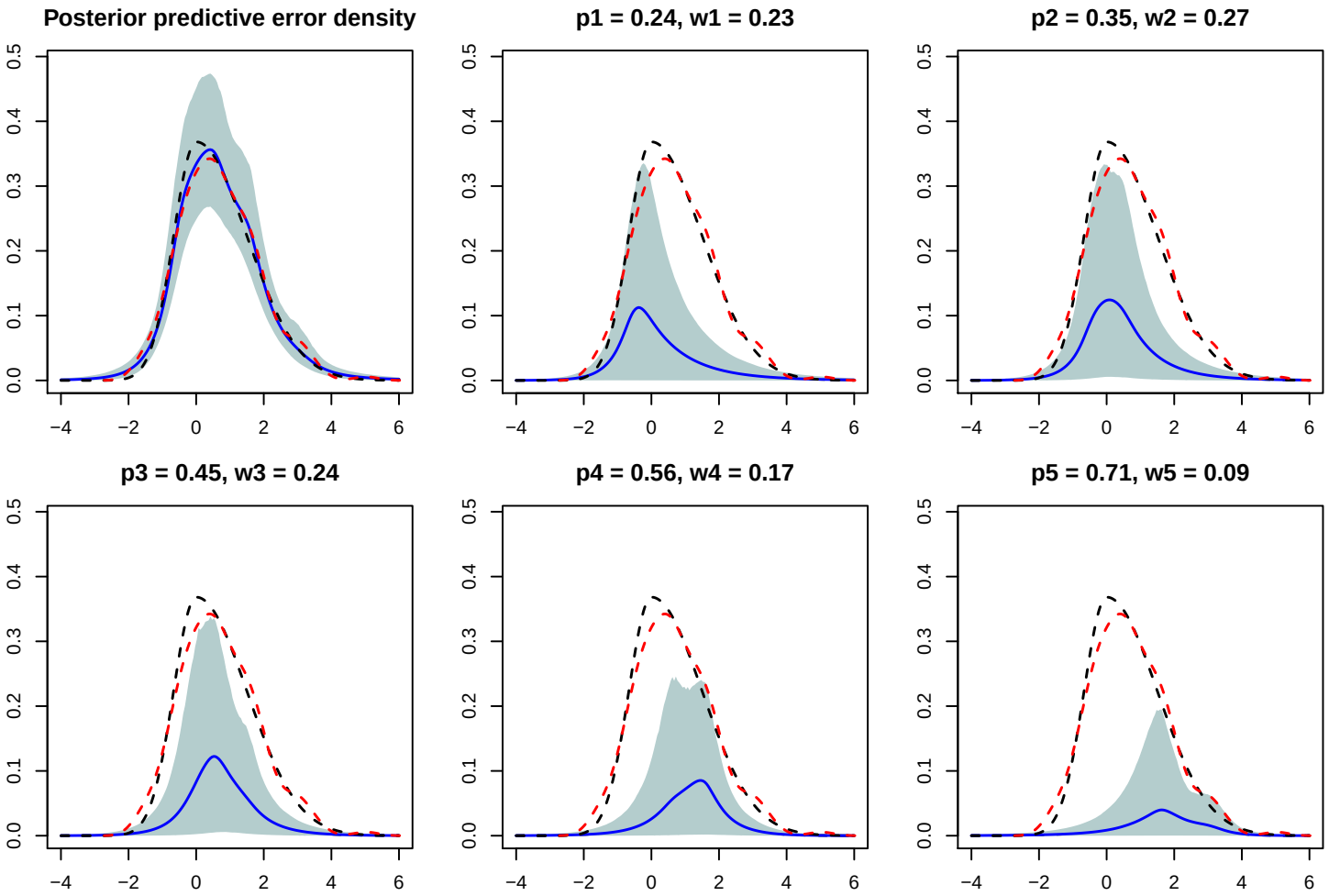


Figure: Synthetic data from skew-normal distribution. Posterior mean and 95% interval estimates for the error density (top left panel) and for the error density weighted components, under the AL mixture model with $K = 5$ and random p_k.

Boston housing data example

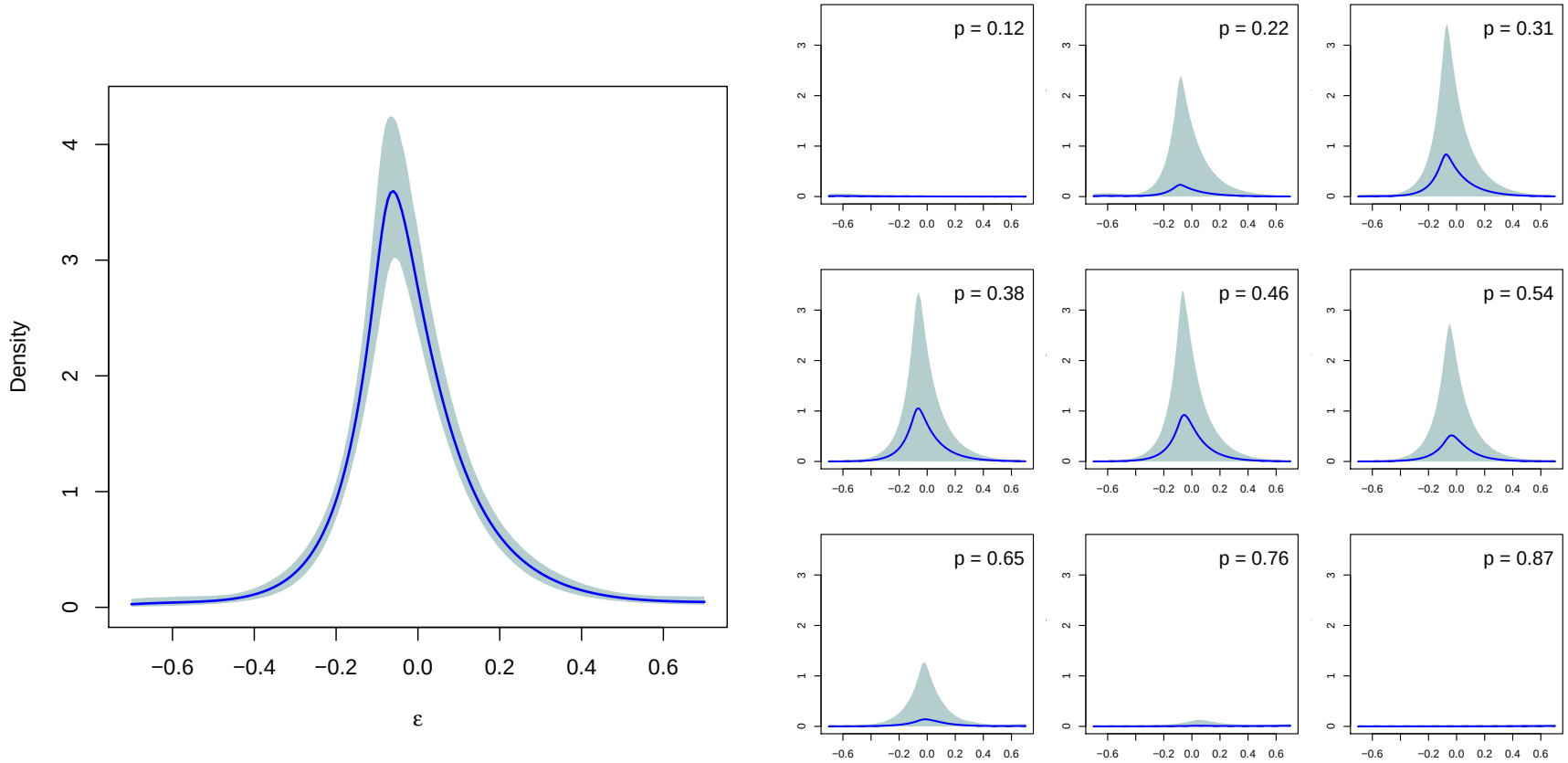
- Realty price data from the Boston area with $n = 506$ observations:
  - response: log-transformed median value of owner-occupied housing in USD 1000;
  - 15 predictors, including: per capita crime (CRIM), nitric oxides concentration (parts per 10 million) per town (NOX), average number of rooms per dwelling (RM), index of accessibility to radial highways per town (RAD), and full-value property-tax rate per USD 10,000 per town (TAX), transformed African American population proportion (B), and percentage values of lower status population (LSTAT).
- Similar inferences under the random-p_k AL mixture and fixed-p_k GAL mixture models (both with $K = 9$).
- Based on posterior predictive criteria (LPML), both models outperform single-quantile regression models (with GAL errors) for essentially any fixed quantile level.

Boston housing data example

- Realty price data from the Boston area with $n = 506$ observations:
  - response: log-transformed median value of owner-occupied housing in USD 1000;
  - 15 predictors, including: per capita crime (CRIM), nitric oxides concentration (parts per 10 million) per town (NOX), average number of rooms per dwelling (RM), index of accessibility to radial highways per town (RAD), and full-value property-tax rate per USD 10,000 per town (TAX), transformed African American population proportion (B), and percentage values of lower status population (LSTAT).
- Similar inferences under the random-p_k AL mixture and fixed-p_k GAL mixture models (both with $K = 9$).
- Based on posterior predictive criteria (LPML), both models outperform single-quantile regression models (with GAL errors) for essentially any fixed quantile level.

Boston housing data example

- Realty price data from the Boston area with $n = 506$ observations:
  - response: log-transformed median value of owner-occupied housing in USD 1000;
  - 15 predictors, including: per capita crime (CRIM), nitric oxides concentration (parts per 10 million) per town (NOX), average number of rooms per dwelling (RM), index of accessibility to radial highways per town (RAD), and full-value property-tax rate per USD 10,000 per town (TAX), transformed African American population proportion (B), and percentage values of lower status population (LSTAT).
- Similar inferences under the random-p_k AL mixture and fixed-p_k GAL mixture models (both with $K = 9$).
- Based on posterior predictive criteria (LPML), both models outperform single-quantile regression models (with GAL errors) for essentially any fixed quantile level.



(a) Error density (b) Weighted mixture components

Figure: Boston housing data. Under the random-p_k AL mixture model, posterior mean and 95% interval estimates for: (a) the error density; (b) the error density weighted components.

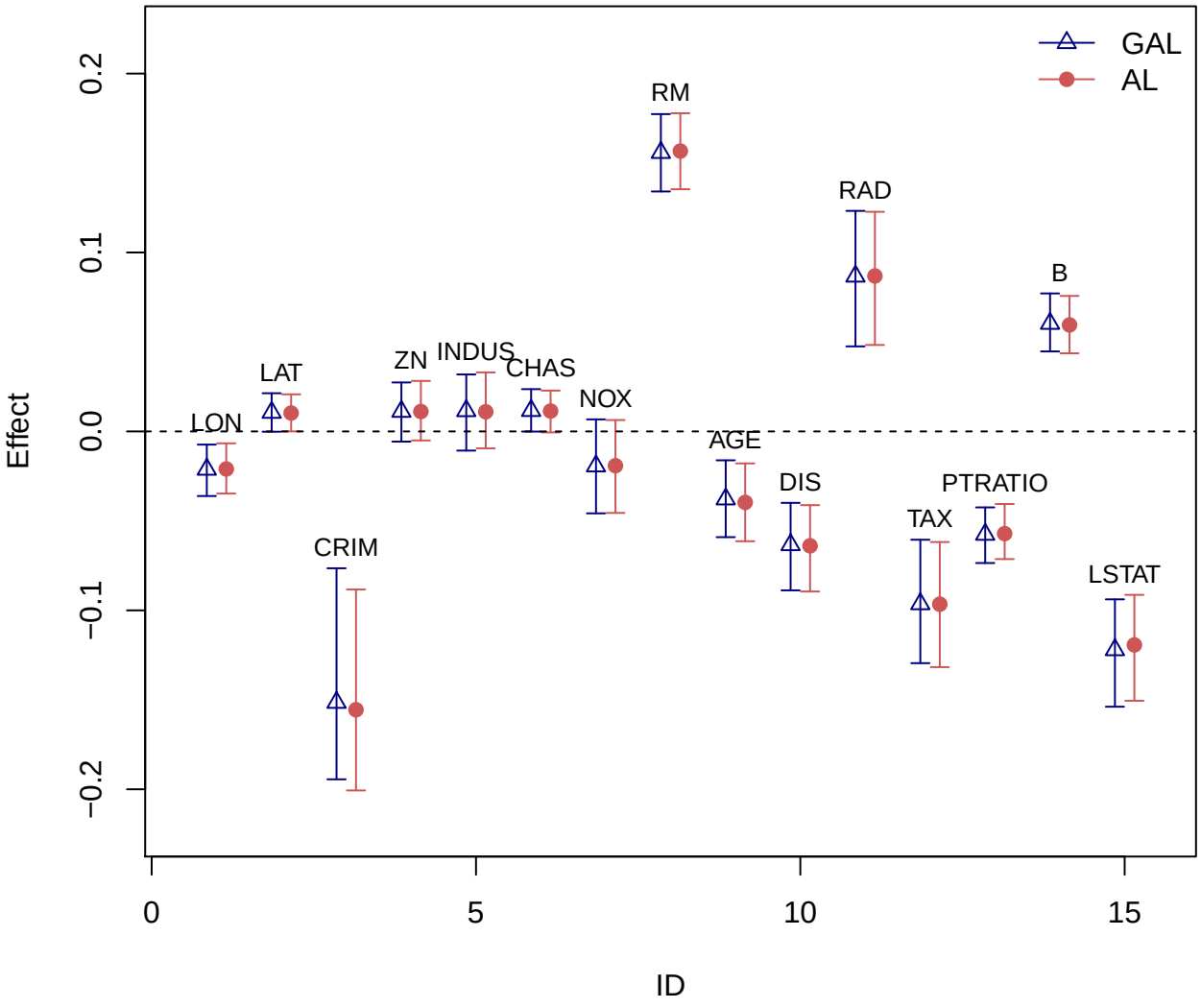


Figure: Boston housing data. Posterior mean and 95% interval for β_j , $j = 1, \ldots, 15$, under the fixed-p_k GAL mixture model, and the random-p_k AL mixture model.

Summary

- A mixture model to:
  - integrate information from multiple parts of the response distribution to inform estimation of covariate effects;
  - identify the most relevant parts of the response distribution through mixture weights associated with different quantiles.
- Two modeling scenarios: mixtures of GAL/AL distributions with fixed/random quantile levels.
- Applications to ROC estimation with covariates, extending the mixture model to incorporate stochastic ordering for the response distributions associated with the infected and non-infected groups.

Summary

- A mixture model to:
  - integrate information from multiple parts of the response distribution to inform estimation of covariate effects;
  - identify the most relevant parts of the response distribution through mixture weights associated with different quantiles.
- Two modeling scenarios: mixtures of GAL/AL distributions with fixed/random quantile levels.
- Applications to ROC estimation with covariates, extending the mixture model to incorporate stochastic ordering for the response distributions associated with the infected and non-infected groups.

Summary

- A mixture model to:
  - integrate information from multiple parts of the response distribution to inform estimation of covariate effects;
  - identify the most relevant parts of the response distribution through mixture weights associated with different quantiles.
- Two modeling scenarios: mixtures of GAL/AL distributions with fixed/random quantile levels.
- Applications to ROC estimation with covariates, extending the mixture model to incorporate stochastic ordering for the response distributions associated with the infected and non-infected groups.

- Acknowledgment: funding from NSF under award SES 1631963.

MANY THANKS!