النموذج الهرمي والتوزيع الأولي للقوى في انحدار التقسيم البيزي

أ. م. د. طاهر ريسان دخل
قسم الإحصاء/ كلية الإدارة والاقتصاد
جامعة القادسية
tahir_reisan@yahoo.com

المستخلص

تم التحري في هذا البحث عن الصلة بين النماذج الهرمية والتوزيع الأولي للقوى في الانحدار التقسيم. بمعلومية مستوى التقسيم، تم تطوير صيغة محددة لمعلمة القوة لمزاولة التوزيع الأولي للقوى للانحدار التقسيم مع النموذج الهرمي المماثل. إضافة إلى ذلك فقد قدرنا العلاقة بين مستوى التقسيم من خلال النموذج الهرمي. وقد تم توضيح المنهجية المقترحة من خلال استخدام بيانات حقيقية.

التصنيفات الرئيسية للبحث: النموذج الهرمي، التوزيع اللاحق، التوزيع الأولي للقوى، الانحدار التقسيم.
1 Introduction

Bayesian parameter estimation in quantile regression (QReg) is often a difficult issue because of a standered conjugate prior distribution is not available. To solve this problem, Alhamzawi and Yu 2011 extended the power prior distribution of Ibrahim and Chen (2000) for Bayesian quantile regression. This prior is a conjugate prior distribution for Bayesian quantile regression. In this paper, we examine the relation between the power prior and the hierarchical model in (QReg). We investigate the relation between the power parameter and the quantile level via the hierarchical model.

(QReg) models have received considerable attention over the years (see, Koenker 2005; Yu et al. 2003; Cade et al 2003). Since Yu and Moyeed (2001) Bayesian inference quantile regression (BQReg) has attracted a lot of attention in literature (see, Hanson and Johnson 2002; Geraci and Bottai 2007; Yu and Stander 2007; Reed and Yu 2009; Lancaster and Jun 2010; Yuan and Yin 2010; Alhamzawi et al. 2011, Kozumi and Kobayashi 2011, Alhamzawi and Yu 2012, Alhamzawi and Yu 2013). However, the prior distribution plays the most important role in Bayesian quantile regression (BQReg). Since being introduced in Ibrahim and Chen (2000), the power prior distribution has become a popular technique to incorporate the historical data into the current data. This power prior distribution has been widely used for a variety of applications. The relation between the power prior distribution and hierarchical models in generalized linear models has been discussed by Chen and Ibrahim (2006). The authors found expressions for the power parameter to calibrate the power prior distribution to a corresponding hierarchical model.

Alhamzawi and Yu 2011 extended the power prior distribution of Ibrahim and Chen (2000) for Bayesian quantile regression.

The rest of this paper is organized as follows. In Section 2, we introduce the hierarchical model in (QReg) based on the mixture representation of the asymmetric Laplace distribution. In Section 3, we define the power prior distribution in (QReg) and we define the power prior distribution based on the mixture representation. In Section 4, we explain the behavior of the posterior under the power prior distribution in (QReg). In Section 5, we present the propriety of the power prior distribution in (QReg), the relation between the power prior distribution and the hierarchical model, and the relation between the power parameter and the quantile. In Section 6, we demonstrate the proposed methodology for obtaining the guide value for the power parameter with real data.
2- Hierarchical Model
Consider the regression model,
\[ y_i = x_i' \beta_p + \varepsilon_i \]  
(1)
where \( y_i \) is the outcome variable, \( x' = (x_{i1}, x_{i2}, \ldots, x_{ik}) \) represent the \( k \) independent variables, \( \beta_p \) is a \( k \times 1 \) vector of regression coefficients and \( \varepsilon_i \), \( i=1, \ldots, n \) represent error terms which are identical and independent distributions. The distribution of the error is assumed unknown and is restricted to have the \( p \)th quantile equal to zero and \( 0 < p < 1 \).

Following Yu and Moyeed (2001), we consider \( \varepsilon \) has asymmetric Laplace distribution (ALD) with density
\[ f(y_i | \beta_p) = p(1-p) \exp\{-p(y_i - x_i' \beta_p)\} \]  
(2)
Where \( p \) determines the quantile level and \( \rho_p(u) = (p - 1(u < 0))u \).

As provided in Reed and Yu (2009) and Kozumi and Kobayashi (2009) that any variable has asymmetric Laplace distribution (ALD) with density (2) can be viewed as a mixture of an exponential and a scaled normal distribution given by
\[ \varepsilon = d(1 - 2p) + \sqrt{2d} \xi, \]  
(3)
Where \( \nu = [p(1-p)]^{-1} \) is a standard exponential variable, then it follows that each \( \nu_i \) has exponential distribution, \( \exp(p(1-p)) \), and \( \xi_i \) is a standard normal distribution. Now, the conditional distribution of each \( y_i \) given \( \nu_i \) is normal with mean \( x_i' \beta_p + (1 - 2p)\nu_i \) and variance \( 2\nu_i \). Thus, the posterior density of \( \beta_p \) is given by,
\[ f(\beta_p | y_i, \nu_i) \propto (\nu_i)^{-\frac{1}{2}} \exp\left\{-\frac{(y_i - x_i' \beta_p)^2}{4\nu_i}\right\}, \]  
(4)
and the complete data density of \( (y_i, \nu_i) \) is then given by
\[ f(y_i, \nu_i | \beta_p) \propto (\nu_i)^{-\frac{1}{2}} \exp\left\{-\frac{(y_i - x_i' \beta_p)^2}{4\nu_i}\right\}\exp\{-p(1-p)\nu_i\}. \]  
(5)
Let \( y = (y_1, \ldots, y_n) \) and \( \nu = (\nu_1, \ldots, \nu_n) \), then the joint density of \( (y, \nu) \) is given by
\[ f(y, \nu | \beta_p) = f(y | \beta_p, \nu) \pi(\nu). \]
If we integrating out the exponential variable, this leads to the likelihood
\[ f(y | \beta_p) = \int f(y | \beta_p, \nu) \pi(\nu) d\nu. \]  
(6)
The model (1) with one historical dataset exist can be written as
\[ y_i = x_i \beta_p + \varepsilon_i, \quad i = 1, \ldots, n, \text{ and } y_{0i} = x_{0i} \beta_{0p} + \varepsilon_{0i}, \quad i = 1, \ldots, n_0, \]  
(7)
where \( \beta_p \) and \( \beta_{0p} \) denote the \( k \) regression coefficients for the current and historical study, respectively, \( x_i \) and \( x_{0i} \) represent the \( k \) known covariates for the current and historical data, respectively, \( \varepsilon_i \) and \( \varepsilon_{0i} \) denote the error term associated with the subject \( i \) for the current and historical study, respectively.

Then we have the following hierarchical model
\[ y_i = x_i \beta_p + (1 - 2p)v_i + \sqrt{2v_i} \xi_i \quad \text{and} \quad y_{0i} = x_{0i} \beta_{0p} + (1 - 2p)v_{0i} + \sqrt{2v_{0i}} \xi_{0i} \]
\[ \beta_p | \mu_0, B_0 \sim N_k(\mu_0, B_0), \quad \beta_{0p} | \mu_0, B_0 \sim N_k(\mu_0, B_0), \]
\[ v_i \sim p(1 - p) \exp\{-p(1 - p)v_i\}, \quad v_{0i} \sim p(1 - p) \exp\{-p(1 - p)v_{0i}\}, \]
\[ \xi_i \sim \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \xi_i^2\right), \quad \xi_{0i} \sim \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \xi_{0i}^2\right), \]  
(8)
where \( \mu_0 \) and \( B_0 \) are known fixed parameters, \( v_{0i} = [p(1 - p)]^{-1} z_{0i}, z_{0i} \) is a standard exponential latent variable for the historical data, \( \xi_{0i} \) is a standard normal variable, and \( z_{0i} \) and \( \xi_{0i} \) are mutually independent.

3 - Power prior
Alhamzawi and Yu (2012) follow Ibrahim and Chen (2000) and define the power prior distribution for \( \beta_p \) in (QReg) for the current study as (Alhamzawi and Yu, 2012)
\[ \pi(\beta_p | D_0, a_0) \equiv \frac{L(\beta_p | D_0)^{a_0} \pi_0(\beta_p)}{\int L(\beta_p | D_0)^{a_0} \pi_0(\beta_p) d\beta_p} \]
\[ = \frac{L(\beta_p | D_0)^{a_0} \pi_0(\beta_p)}{g(a_0)} \]
\[ \propto L(\beta_p | D_0)^{a_0} \pi_0(\beta_p) \]
\[ = \left[ \prod_{i=1}^{n_0} \beta_p^{a_0} (1 - \beta_p)^{a_0} \exp\{-a_0 \beta_p (y_{0i} - x_{0i} \beta_p)\} \right] \pi_0(\beta_p) \]  
(9)
where $D_0$ represents the historical data, $0 \leq \alpha_0 \leq 1$, $\alpha_0$ determines by expert opinion, $L(\beta_p | D_0)$ denotes the likelihood function, $\pi_0(\beta_p)$ denotes the initial prior for $\beta_p$. Under the mixture representation (3), the joint power prior distribution for $\beta_p$ and $V_0$ is given by,

$$\pi(\beta_p, v_0 | D_0, \alpha_0) \propto \left[ \prod_{i=1}^{\infty} f(\beta_p | Y_{0i}, v_{0i}) \right]^{\alpha_0} \pi_0(\beta_p),$$

(10)

where $f(\beta_p | Y_{0i}, v_{0i})$ is (4) with $(Y_{0i}, v_{0i})$ in place of $(y_i, v_i)$, $v_0 = (v_{01}, ..., v_{0n})'$ and $v_{0i}$ has exponential distribution, Exp(p(1− p)). The power priors (9) and (10) have several attractive properties. First, the power prior (9) can be obtained from the power prior (10) by integrating out the exponential variable. Further, the power priors (9) and (10) are always proper and have lower and upper bounds even if $\pi_0(\beta_p)$ is improper. In addition, the priors (9) and (10) depend on the quantile. The prior specification is completed by specifying a prior distribution for $\beta_p$. Let $D$ denote to the current data and $V = (v_1, ..., v_n)'$, then the joint posterior distribution of $\beta_p, v$ and $v_0$ is given by (Alhamzawi and Yu, 2012)

$$\pi(\beta_p, v, v_0 | D, D_0, \alpha_0) \propto \left[ \prod_{i=1}^{n} f(\beta_p | Y_i, v_i) \right]^{\alpha_0} \pi_0(\beta_p),$$

(11)

4. Posterior Behavior under the power prior

To demonstrate the behavior of the marginal posterior distribution of $\beta_p$ under the power prior with respect to different values for the power parameter. We simulate two data sets for the current and historical study. For the current study, 350 observation was generate from the model $Y_i = 10 - x_i + \varepsilon_i$ where $x_i$ was simulated from a uniform distribution on the interval (0, 10) and $\varepsilon_i \sim N(0, 1)$.

For the historical data, 200 observation was generate from the model $Y_{0i} = 9 - 1.5x_{0i} + \varepsilon_{0i}$, where $x_{0i}$ was simulated from a uniform distribution on the interval (0, 10) and $\varepsilon_{0i} \sim N(0, 1)$. We take a multivariate normal distribution with mean zero and variance covariance matrix $B = 100I$ as initial prior for $\beta_p$. Figure 1 and 2 compared the marginal posterior densities for $\beta_{(0)}p$ and $\beta_{(1)}p$ for $p=95\%$ and $75\%$ respectively m for improper prior with the posterior densities of $\beta_{(0)}p$ and $\beta_{(1)}p$ for the power prior. Clearly, the power prior is more informative than improper prior, due to the small range of posterior densities.
Figure 1: plots of posterior densities for $\beta_{0.95}$ where the dotted curve is for improper uniform prior ($a_0=0$), the dashed and solid curves are for power priors with power parameter $a_0=0.50$ and 0.90 respectively.

Figure 2: plots of posterior densities for $\beta_{0.75}$ where the dotted curve is for improper uniform prior ($a_0=0$), the dashed and solid curves are for power priors with power parameter $a_0=0.50$ and 0.90 respectively.
5. Main results

The power prior proposed by Ibrahim and Chen (2000) has been constructed to be a useful class of informative prior in Bayesian analysis. This prior depend on the availability of the historical data, and in the context of Bayesian analysis when such data is available the prior would be better proper because it is well known that any informative Bayesian analysis requires a proper prior distribution, thus the propriety of the power prior is of critical importance. In this section we discuss the relation between power priors and hierarchical models. Thus, we present Lemma 1 and 2 to introduce the marginal posterior distribution for $\beta_p$ under the power prior, Lemma 3 introduce the marginal posterior distribution for $\beta_p$ under the hierarchical model and Lemma 4 discuss the relation between power priors and hierarchical models.

Lemma 1. The marginal posterior distribution of $\beta_p$ under the power prior (10) with multivariate normal distribution as initial prior for $\beta_p$ is given by

$$
\beta_p | y, X, V, y_0, X_0, V_0, a_0 \sim N_k(A^{-1}B, A^{-1}),
$$

Where

$$
A = \frac{1}{2} X'VX + \frac{a_0}{2} X_0'V_0X_0 + B_0^{-1},
B = \frac{1}{2} X'Vu + \frac{a_0}{2} X_0'V_0u_0 + B_0^{-1}\mu_0,
$$

here, $y = (y_1, ..., y_n)$, $X = (x_1, ..., x_n)'$, $y_0 = (y_{01}, ..., y_{0n_0})'$, $X_0 = (x_{01}, ..., x_{0n_0})'$,

$$
V = diag(v_1, ..., v_n), \ V_0 = diag(v_{01}, ..., v_{0n_0}), \ u = (u_1, ..., u_n), \ u_0 = (u_{01}, ..., u_{0n_0}),
$$

with $u_i = y_i - (1 - 2p)v_i$ and $u_{0i} = y_{0i} - (1 - 2p)v_{0i}$.

The proof of Lemma (1) with the details of the Gibbs sampler is given in Appendix.

Lemma 2. The marginal posterior distribution of $\beta_p$ under the power prior (10) with uniform prior distribution as initial prior for $\beta_p$ is given by

$$
\beta_p | y, X, V, y_0, X_0, V_0, a_0 \sim N_k(A_1^{-1}B_1, A_1^{-1}),
$$

Where

$$
A_1 = \frac{1}{2} X'VX + \frac{a_0}{2} X_0'V_0X_0,
B_1 = \frac{1}{2} X'Vu + \frac{a_0}{2} X_0'V_0u_0,
$$

The proof of Lemma (2) is similar to proof of Lemma (1).

Lemma 3. The marginal posterior distribution of $\beta_p$ for the hierarchical model (8) in (QReg) is given by
\( \beta_p | y, X, V, y_0, X_0, V_0 \sim N_k (A_2^{-1}B_2, A_2^{-1}) \),

Where
\[ A_2 = \frac{1}{2}X'X + B_0^{-1} - (2B_0 - \left( B_0^{-1} + \frac{1}{2}X_0'V_0X_0 \right)^{-1})^{-1}, \]
\[ B_2 = \frac{1}{2}X'V_0 + \left( 2B_0 - \left( B_0^{-1} + \frac{1}{2}X_0'V_0X_0 \right)^{-1} \right)^{-1} \left( B_0^{-1} + \frac{1}{2}X_0'V_0X_0 \right)^{-1} \left( \frac{1}{2}X_0'V_0u_0 \right). \]

The proof of Lemma 3 in Appendix.

**Lemma 4.** The posterior distributions of the quantile coefficient \( \beta_p \) given in Lemma 2 and 3 are identical distributions if and only if
\[ a_0 (I + B_0X_0V_0X_0) = I \quad (14) \]

**Proof:** Similar to Chen and Ibrahim (2006), If \( A_2 = A_2 \) then we have
\[ \frac{1}{2}X'X + \frac{a_0}{2}X_0'V_0X_0 = \frac{1}{2}X'X + B_0^{-1} - (2B_0 - \left( B_0^{-1} + \frac{1}{2}X_0'V_0X_0 \right)^{-1})^{-1}, \]
and this lead to
\[ 2^{-1}a_0B_0X_0'V_0X_0 = I - (2I - (I + 2^{-1}B_0X_0'V_0X_0^{-1})^{-1}). \]

A little algebra shows
\[ a_0B_0X_0'V_0X_0 \left[ B_0X_0'V_0X_0 + I \right] = B_0X_0'V_0X_0 \]
\[ a_0 (I + B_0X_0'V_0X_0) = I \quad (15) \]

Similarly, it can be shown that \( B_2 = B_2 \) if and only if \( a_0[B_0X_0'V_0X_0 + I] = I \).

Like Chen and Ibrahim (2006), we can use the connection between the power prior and the hierarchical model to specify a guide value for \( a_0 \) in (QReg). To achieve this we can write equation (15) as
\[ a_0[B_0X_0'Z_0X_0 + p(1 - p)] = p(1 - p), \]
where \( Z_0 = d\text{ia}(z_{01}, ..., z_{0n}) \).

Since \( Z_0 \) is random then the guide value for \( a_0 \) is the posterior expectation of
\[ k\rho(1 - p) \]
\[ k\rho(1 - p) + tr(B_0X_0'Z_0X_0) \]
That is,
\[ \hat{a}_0 = E \left( \frac{k\rho(1 - p)}{k\rho(1 - p) + tr(E_0X_0'E_0X_0)} \right) \quad (16) \]

The posterior expectation is taken with respect to \( Z_0 \), where \( B_0 \) is constant.

Equation (16) reflects the relation between the power parameter and the quantile level.
6. Wage data

We consider data from the British Household Panel Survey. The data represent the wage distribution among British workers which was previously analyzed by Yu et al. (2005) and Alhamzawi and Yu (2012). Similar to Alhamzawi and Yu (2012), we use the data for the year 2000 as current data and for year 1994 as historical data. Our model is described as follows

\[ \ln(Y_i) = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 D_i, \]

where \( S_i \) is the number of years of schooling, \( E_i \) is the potential experience (approximated by the age minus years of schooling minus 6), and \( D_i \) is equal to 1 for public sector workers and 0 otherwise. We consider (QReg) model to fit the current and the historical data. We take \( B_0 \) to be a fixed diagonal matrix such that \( B_0 = 100I \). We use Gibbs sampler to sample \( \beta_p \) and \( \beta_{0p} \) from their respective distribution. We specify \( \alpha_0 \) from equation (16). Table 1 summarizes the posterior mean for \( \beta_p \) and \( \beta_{0p} \) under the hierarchical model. The posterior distribution for \( \beta_p \) under the power prior are summarized in Table 2 for different value for the power parameter including \( \alpha_0 \). Clearly, the posterior distribution for the regression parameter under the power prior with \( \alpha_0 \) are fairly close to those obtained under the hierarchical model.

7. Appendixes

Proof Lemma 1. First, consider equation (4). In vector notation, the likelihood function for current and historical data are, respectively, given by

\[ f(y|\beta_p, \sigma, v) \propto \sigma^{-\frac{n}{2}} (\prod_{i=1}^{n} v_i^{-\frac{1}{2}}) \exp \left( -\frac{1}{4}(u - X\beta_p)^\top V(u - X\beta_p) \right) \]

\[ f(y_0|\beta_p, \sigma, v_0) \propto \sigma^{-\frac{n_0}{2}} (\prod_{i=1}^{n_0} v_{0i}^{-\frac{1}{2}}) \exp \left( -\frac{1}{4}(u_0 - X_0\beta_p)^\top V_0(u_0 - X_0\beta_p) \right) \]

Here, \( y = (y_1, ..., y_n) \), \( x = (x_1, ..., x_n) \), \( y_0 = (y_{01}, ..., y_{0n_0}) \), \( x_0 = (x_{01}, ..., x_{0n_0}) \), \( V = \text{diag}(v_1, ..., v_n) \), \( V_0 = \text{diag}(v_{01}, ..., v_{0n_0}) \), \( u = (u_1, ..., u_n) \), \( u_0 = (u_{01}, ..., u_{0n_0}) \), \( u_i = y_i - (1 - 2p)v_i \), and \( u_{0i} = y_{0i} - (1 - 2p)v_{0i} \).

The posterior distribution of \( \beta_p \) can then be calculated using the joint posterior distribution in equation (11). We have

\[ f(\beta_p, v, v_0|\sigma, \sigma_0, \alpha_0) \propto (\prod_{i=1}^{n} v_i^{-\frac{1}{2}}) \exp \left( -\frac{1}{4}(u - X\beta_p)^\top V(u - X\beta_p) \right) \times \prod_{i=1}^{n} \exp \{-p(1 - p)v_i\} \]

\[ \times (\prod_{i=1}^{n_0} v_{0i}^{-\frac{1}{2}}) \exp \left( -\frac{\sigma_0^2}{4}(u_0 - X_0\beta_p)^\top V_0(u_0 - X_0\beta_p) \right) \]
From (19), the full conditional distribution of $\beta_p$ is given by
\[ f(\beta_p | \sigma, v, v_0, D, D_0, \alpha_0) \propto \exp\left\{-\frac{1}{4} (u - X\beta_p)'V(u - X\beta_p) - \frac{\alpha_0}{4} (u_0 - X_0\beta_p)'V_0(u_0 - X_0\beta_p)\right\} \]
\[ \times \exp\left\{-\frac{1}{2} (\beta_p - \mu_0)'B_0^{-1}(\beta_p - \mu_0)\right\} \]  
(19)

We have $(u - X\beta_p)'V(u - X\beta_p)$ and $\alpha_0(u_0 - X_0\beta_p)'V_0(u_0 - X_0\beta_p)$ into sum of squares
\[ (u - X\beta_p)'V(u - X\beta_p) = (u - X\beta_p)'V(u - X\beta_p) + (\beta_p - \beta_p^*)'X'VX(\beta_p - \beta_p^*). \]
\[ \alpha_0(u_0 - X_0\beta_p)'V_0(u_0 - X_0\beta_p) = \alpha_0((u_0 - X_0\beta_p)'V_0(u_0 - X_0\beta_p) + (\beta_p - \beta_p^*)'X_0'V_0X_0(\beta_p - \beta_p^*). \]

We set $X'VX\beta_p = X'V$, and $X_0'V_0X_0\beta_p = X_0'V_0u_0$. Then the posterior distribution of $\beta_p$ is given by
\[ \beta_p | y, X, V, y_0, X_0, V_0, \alpha_0 \sim N_k(A^{-1}B, A^{-1}). \]

Where
\[ A = \frac{1}{2} X'VX + \frac{\alpha_0}{2} X_0'V_0X_0 + B_0^{-1}, \]
\[ B = \frac{1}{2} X'V_0 + \frac{\alpha_0}{2} X_0'V_0u_0 + B_0^{-1}\mu_0. \]

To complete our MCMC-based computation technique, the full conditional distribution of $v_{0i}$ is given by
\[ \pi(v_{0i} | \beta_p, \alpha_0, D_0) \propto (v_{0i})^{-\alpha_0/2} \exp\left\{-\frac{\alpha_0}{4v_{0i}} (y_{0i} - x_{0i}^T\beta_p - (1 - 2p)v_{0i})^2 - p(1 - p)v_{0i}\right\} \]
\[ \propto (v_{0i})^{-\alpha_0/2} \exp\left\{-\frac{\alpha_0}{4v_{0i}} (y_{0i} - x_{0i}^T\beta_p)^2 - \left(\frac{\alpha_0(1 - 2p)^2}{4}\right) + p(1 - p)v_{0i}\right\} \]
\[ = v_{0i}^{-\alpha_0/2} \exp\left\{-\frac{1}{2} \left[\frac{\alpha_0(y_{0i} - x_{0i}^T\beta_p)^2}{2v_{0i}} + \left(\frac{\alpha_0(1 - 2p)^2}{4} + 2p(1 - p)v_{0i}\right)\right] \right\} \]

Thus, the full conditional distribution of $v_{0i}$ is a generalized inverse Gaussian (GIG) distribution. In the same way we can deduce that the full conditional distribution of $v_{0i}$ is also GIG distribution; that is,
\[ \pi(v_i | \beta_p) \propto v_i^{-1/2} \exp\left\{-\frac{1}{2} \left[\frac{(y_i - x_i^T\beta_p)^2}{2v_i} + \left(\frac{1}{2}\right)v_i\right] \right\} \]
Proof of Lemma 3.

\[ \pi(\beta_p, \beta_0, \mu_0 | y, y_0, v, v_0) \propto \exp \left\{ -\frac{1}{4} (u - X \beta_p)'V(u - X \beta_p) \exp \left\{ -\frac{1}{4} (u_0 - X_0 \beta_p)'V_0(u_0 - X_0 \beta_p) \right\} \right. \\
\times \exp \left\{ -\frac{1}{2} (\beta_p - \mu_0)'B_0^{-1}(\beta_p - \mu_0) \right\} \exp \left\{ -\frac{1}{2} (\beta_p - \mu_0)'B_0^{-1}(\beta_p - \mu_0) \right\}. \]

After integrating out \( \beta_0 \) and rearrange the terms, we can get

\[ \pi(\beta_p, \mu_0 | y, y_0, v, v_0) \propto \exp \left\{ -\frac{1}{4} (u - X \beta_p)'V(u - X \beta_p) - \frac{1}{2} (\beta_p - \mu_0)'B_0^{-1}(\beta_p - \mu_0) \right\} \exp \left\{ -\frac{1}{2} \left[ \mu_0 \left( B_0^{-1} - \left( B_0 + \frac{1}{2} B_0 X_0' V_0 X_0 B_0^{-1} \right) \right)^{-1} \mu_0 \right] \left[ I + \frac{1}{2} X_0' V_0 X_0 B_0^{-1} \right] \left( X_0' V_0 u_0 \right) \right\}. \]

Then, integrating out \( \mu_0 \) leads to

\[ \pi(\beta_p | y, y_0, v, v_0) \propto \left\{ -\frac{1}{2} \left[ \beta_p' \left( \frac{1}{2} X' V X + B_0^{-1} - \left( 2B_0 - \left( B_0^{-1} + \frac{1}{2} X_0' V_0 X_0 \right)^{-1} \right) \right)^{-1} \right] \right\} \beta_p \]

\[ -\beta_p' X' V u + \left( 2B_0 - \left( B_0^{-1} + \frac{1}{2} X_0' V_0 X_0 \right)^{-1} \right)^{-1} \left( B_0^{-1} + \frac{1}{2} X_0' V_0 X_0 \right)^{-1} \left( X_0' V_0 u_0 \right). \]

Then we have

\( \beta_p | y, X, V, y_0, X_0, V_0 \sim N_k( A_2^{-1} B_2, A_2^{-1}) \).

Where

\[ A_2 = \frac{1}{2} X' V X + B_0^{-1} - \left( 2B_0 - \left( B_0^{-1} + \frac{1}{2} X_0' V_0 X_0 \right)^{-1} \right)^{-1}, \]

\[ B_2 = \frac{1}{2} X' V u + \left( 2B_0 - \left( B_0^{-1} + \frac{1}{2} X_0' V_0 X_0 \right)^{-1} \right)^{-1} \left( B_0^{-1} + \frac{1}{2} X_0' V_0 X_0 \right)^{-1} \left( B_0^{-1} + \frac{1}{2} X_0' V_0 X_0 \right)^{-1} \left( B_0^{-1} + \frac{1}{2} X_0' V_0 X_0 \right)^{-1} \left( X_0' V_0 u_0 \right). \]
Table 1: Posterior estimates of $\beta_{0(p)}$ and $\beta_{00(p)}$ under the hierarchical model.

<table>
<thead>
<tr>
<th>$p$</th>
<th>parameter</th>
<th>posterior mean</th>
<th>parameter</th>
<th>posterior mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>$\beta_{0(p)}$</td>
<td>7.163</td>
<td>$\beta_{00(p)}$</td>
<td>7.188</td>
</tr>
<tr>
<td></td>
<td>$\beta_{1(p)}$</td>
<td>0.027</td>
<td>$\beta_{01(p)}$</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>$\beta_{2(p)}$</td>
<td>0.004</td>
<td>$\beta_{02(p)}$</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>$\beta_{3(p)}$</td>
<td>-0.103</td>
<td>$\beta_{03(p)}$</td>
<td>-0.125</td>
</tr>
<tr>
<td>0.75</td>
<td>$\beta_{0(p)}$</td>
<td>6.840</td>
<td>$\beta_{00(p)}$</td>
<td>6.798</td>
</tr>
<tr>
<td></td>
<td>$\beta_{1(p)}$</td>
<td>0.015</td>
<td>$\beta_{01(p)}$</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>$\beta_{2(p)}$</td>
<td>0.009</td>
<td>$\beta_{02(p)}$</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>$\beta_{3(p)}$</td>
<td>-0.031</td>
<td>$\beta_{03(p)}$</td>
<td>-0.028</td>
</tr>
<tr>
<td>0.50</td>
<td>$\beta_{0(p)}$</td>
<td>6.799</td>
<td>$\beta_{00(p)}$</td>
<td>6.806</td>
</tr>
<tr>
<td></td>
<td>$\beta_{1(p)}$</td>
<td>0.020</td>
<td>$\beta_{01(p)}$</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>$\beta_{2(p)}$</td>
<td>0.003</td>
<td>$\beta_{02(p)}$</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>$\beta_{3(p)}$</td>
<td>0.012</td>
<td>$\beta_{03(p)}$</td>
<td>0.063</td>
</tr>
<tr>
<td>0.25</td>
<td>$\beta_{0(p)}$</td>
<td>6.572</td>
<td>$\beta_{00(p)}$</td>
<td>6.528</td>
</tr>
<tr>
<td></td>
<td>$\beta_{1(p)}$</td>
<td>0.022</td>
<td>$\beta_{01(p)}$</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>$\beta_{2(p)}$</td>
<td>0.006</td>
<td>$\beta_{02(p)}$</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>$\beta_{3(p)}$</td>
<td>0.066</td>
<td>$\beta_{03(p)}$</td>
<td>0.097</td>
</tr>
<tr>
<td>0.05</td>
<td>$\beta_{0(p)}$</td>
<td>6.334</td>
<td>$\beta_{00(p)}$</td>
<td>6.313</td>
</tr>
<tr>
<td></td>
<td>$\beta_{1(p)}$</td>
<td>0.019</td>
<td>$\beta_{01(p)}$</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>$\beta_{2(p)}$</td>
<td>0.003</td>
<td>$\beta_{02(p)}$</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>$\beta_{3(p)}$</td>
<td>0.098</td>
<td>$\beta_{03(p)}$</td>
<td>0.114</td>
</tr>
</tbody>
</table>
Table 2: Posterior estimates of $\beta_\{(p)\}$ under the power prior distribution.

<table>
<thead>
<tr>
<th>p</th>
<th>$a_0$</th>
<th>$\beta_0(p)\ (95%\ CrI)$</th>
<th>$\beta_1(p)\ (95%\ CrI)$</th>
<th>$\beta_2(p)\ (95%\ CrI)$</th>
<th>$\beta_3(p)\ (95%\ CrI)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>0</td>
<td>6.092 (5.933, 6.147)</td>
<td>0.170 (0.112, 0.226)</td>
<td>0.013 (0.010, 0.033)</td>
<td>-0.347 (-0.834, 0.121)</td>
</tr>
<tr>
<td>0.75</td>
<td>0</td>
<td>7.170 (7.138, 7.205)</td>
<td>0.027 (0.015, 0.043)</td>
<td>0.016 (0.002, 0.027)</td>
<td>-0.119 (-0.138, -0.098)</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>7.403 (7.434, 7.553)</td>
<td>0.031 (0.028, 0.035)</td>
<td>0.011 (0.009, 0.012)</td>
<td>-0.160 (-0.196, -0.147)</td>
</tr>
<tr>
<td>0.75</td>
<td>0</td>
<td>5.941 (5.853, 6.027)</td>
<td>0.071 (0.063, 0.125)</td>
<td>0.011 (0.009, 0.023)</td>
<td>0.027 (0.012, 0.048)</td>
</tr>
<tr>
<td>0.49</td>
<td>0</td>
<td>6.976 (6.923, 7.028)</td>
<td>0.021 (0.019, 0.024)</td>
<td>0.010 (0.008, 0.011)</td>
<td>-0.004 (-0.055, 0.025)</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>7.012 (6.933, 7.071)</td>
<td>0.026 (0.025, 0.031)</td>
<td>0.010 (0.008, 0.011)</td>
<td>-0.030 (-0.056, -0.006)</td>
</tr>
<tr>
<td>0.50</td>
<td>0</td>
<td>5.819 (5.665, 5.962)</td>
<td>0.030 (0.022, 0.054)</td>
<td>0.004 (0.001, 0.013)</td>
<td>0.037 (-0.019, 0.114)</td>
</tr>
<tr>
<td>0.337</td>
<td>0</td>
<td>6.834 (6.702, 6.976)</td>
<td>0.021 (0.015, 0.026)</td>
<td>0.007 (0.001, 0.011)</td>
<td>0.030 (-0.051, 0.116)</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>6.851 (6.693, 6.991)</td>
<td>0.022 (0.016, 0.028)</td>
<td>0.007 (0.003, 0.011)</td>
<td>0.033 (-0.050, 0.113)</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>5.615 (5.535, 5.702)</td>
<td>0.013 (0.003, 0.019)</td>
<td>0.002 (0.001, 0.036)</td>
<td>0.041 (0.005, 0.117)</td>
</tr>
<tr>
<td>0.51</td>
<td>0</td>
<td>6.557 (6.515, 6.600)</td>
<td>0.021 (0.019, 0.023)</td>
<td>0.006 (0.005, 0.007)</td>
<td>0.071 (0.042, 0.104)</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>6.492 (6.457, 6.531)</td>
<td>0.018 (0.016, 0.019)</td>
<td>0.006 (0.005, 0.007)</td>
<td>0.061 (0.032, 0.092)</td>
</tr>
<tr>
<td>0.05</td>
<td>0</td>
<td>5.470 (5.395, 5.546)</td>
<td>0.006 (0.001, 0.013)</td>
<td>-0.005 (-0.003, 0.006)</td>
<td>0.091 (0.035, 0.153)</td>
</tr>
<tr>
<td>0.57</td>
<td>0</td>
<td>6.337 (6.307, 6.365)</td>
<td>0.019 (0.017, 0.021)</td>
<td>0.004 (0.003, 0.005)</td>
<td>0.104 (0.081, 0.127)</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>5.805 (5.850, 5.927)</td>
<td>0.013 (0.012, 0.015)</td>
<td>0.004 (0.003, 0.006)</td>
<td>0.216 (0.181, 0.248)</td>
</tr>
</tbody>
</table>
References


A Note on the Hierarchical Model and Power Prior Distribution in Bayesian Quantile Regression

Abstract.

In this paper, we investigate the connection between the hierarchical models and the power prior distribution in quantile regression (QReg). Under specific quantile, we develop an expression for the power parameter ($a_0$) to calibrate the power prior distribution for quantile regression to a corresponding hierarchical model. In addition, we estimate the relation between the $a_0$ and the quantile level via hierarchical model. Our proposed methodology is illustrated with real data example.

Keywords: Hierarchical model; Posterior distribution; Power prior; Quantile Regression (QReg).