Published in final edited form as:

J Am Stat Assoc. 2011 December 1; 106(496): 1405–1417. doi:10.1198/jasa.2011.tm10470..

Semiparametric Approach to a Random Effects Quantile Regression Model

Mi-Ok Kim[Associate Professor] and

Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center,
Cincinnati, OH 45229-3039

Yunwen Yang, Ph.D.[Student]

Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820

Mi-Ok Kim: miok.kim@cchmc.org; Yunwen Yang: yang31@illinois.edu

Abstract

We consider a random effects quantile regression analysis of clustered data and propose a semiparametric approach using empirical likelihood. The random regression coefficients are assumed independent with a common mean, following parametrically specified distributions. The common mean corresponds to the population-average effects of explanatory variables on the conditional quantile of interest, while the random coefficients represent cluster specific deviations in the covariate effects. We formulate the estimation of the random coefficients as an estimating equations problem and use empirical likelihood to incorporate the parametric likelihood of the random coefficients. A likelihood-like statistical criterion function is yield, which we show is asymptotically concave in a neighborhood of the true parameter value and motivates its maximizer as a natural estimator. We use Markov Chain Monte Carlo (MCMC) samplers in the Bayesian framework, and propose the resulting quasi-posterior mean as an estimator. We show that the proposed estimator of the population-level parameter is asymptotically normal and the estimators of the random coefficients are shrunk toward the population-level parameter in the first order asymptotic sense. These asymptotic results do not require Gaussian random effects, and the empirical likelihood based likelihood-like criterion function is free of parameters related to the error densities. This makes the proposed approach both flexible and computationally simple. We illustrate the methodology with two real data examples.

Keywords

Empirical likelihood; Markov Chain Monte Carlo; Quasi-posterior distribution

1. INTRODUCTION

We extend quantile regression (QR) method to a random effects analysis of clustered data in this paper. A dominant paradigm of clustered data analysis is a Gaussian structure where the random effects and random errors are both assumed identically distributed among themselves, following common Gaussian cumulative distributions respectively. Under this structure, however, the effects of the explanatory variables are assumed to affect only the location of the conditional distribution of the response. Such exclusive attention to the location is constraining as the analysis is often motivated by a natural heterogeneity among the clusters. Moreover the response variable is often analyzed on a transformed scale in practice to meet the rather rigid homoscedastic Gaussian distribution assumptions. The results are, however, less straightforward and difficult to interpret on the original scale. In addition the conditional mean analysis, although most popular, does not necessarily adequately address research questions of interest as shown in the Network for the

Improvement for Addiction Treatment example below. QR can provide an alternative analytic tool. Conditional regression quantiles are invariant under monotonic transformation like quantiles in a univariate case and are modelled locally for their relationships with the explanatory variables with the relationships allowed to differ from one another.

The aforementioned features of QR has been well recognized with independent data. Methodologies are fully developed and the model is widely applied (see Koenker (2005) for an overview). For the analysis of clustered data, however, a limited number of approaches have been proposed. Jung (1996) considered fixed effects median regression and proposed a quasi-likelihood approach. Koenker (2004) considered a random intercept model and proposed a l_1 penalty approach. The l_1 penalty approach is less stringent in its assumptions than most other methods. While the results of the l_1 penalty approach depend on the choice of a penalty parameter, inference of the fixed effects was not studied with an empirically chosen penalty parameter. The method also may not be applicable to more complex random effects model as it essentially treats random effects as parameters and the increasing dimensionality can be an issue. Geraci and Bottai (2007) assumed an asymmetric Laplace error distribution and proposed an expectation-maximization (EM) estimator. Assuming a common asymmetric Laplace distribution, their method constrains the errors to be not only homoscedastic but also have a mode at the median.

In the Bayesian analysis frame work, several parametric approaches have been proposed, similarly using asymmetric Laplace error densities and mostly for the analysis of independent data (e.g., Yu and Moyeed (2001)). Nonparametric approaches were proposed to avoid the restrictive parametric assumption (e.g., Hanson and Johnson (2002), Kottas and Gelfand (2001) and Kottas and Krnjajic (2009)). Although capturing more general forms of skewness and tail behaviors, these nonparametric approaches also restrict the error densities to necessarily have their modes at the quantile of interest. Reich et al. (2010) relaxed this restriction with an infinite mixture of quantile restricted two component Gaussian mixture densities. These nonparametric Bayesian approaches, however, essentially model the error densities, although avoiding parametrically specifying them. As a consequence they can accommodate the error heteroscedasticity only by correctly specifying its form parametrically in the model. This informative modeling requirement of the error heteroscedasticity is restrictive compared with the generality of independent data QR analysis methodology, and make the computation complex.

We propose a semiparametric approach that does not require modeling of the error densities, thereby realizing all the desirable features of the independent data QR analysis without computational complexity. We assume random regression coefficients which may not be necessarily identically distributed or Gaussian. The random coefficients have a common mean which corresponds to the population-average effects of the explanatory variables on the conditional quantile of interest. The random coefficients represent cluster specific deviations in the covariate effects. Under appropriate conditions discussed later the median regression coincides with the conditional mean, and appears as a robust alternative against outliers in the responses. We consider the estimation of the random effects as an estimating equations problem and use empirical likelihood (EL) to incorporate the parametric likelihood of the random effects. We yield a semiparametric likelihood-like criterion function, which we show is asymptotically concave in a neighborhood of the true parameter value and motivates the maximizer as a natural estimator. We use the Bayesian framework and Markov Chain Monte Carlo (MCMC) samplers for the computation.

Monte Carlo methods have been used for classical estimation problems under various settings where classical methods are handicapped by computational difficulties (see Tian et al. (2007) for an overview). Recent works include Chernozhukov and Hong (2003) and Tian

et al. (2007). Chernozhukov and Hong (2003) particularly considered EL and censored QR for the analysis of independent data. In this paper we are concerned with random effects QR analysis and semiparametric likelihood-like criterion function.

A few works in the Bayesian literature also have considered likelihood-like or non-likelihood statistical criterion functions (e.g. Lavine (1995); Dunson et al. (2003); Dunson and Taylor (2005); Lazar (2003), Schennach (2005), Lancaster and Jun (2010)). They were motivated by the computational complexities entailed in nonparametric Bayesian methods and the difficulty of likelihood specification. Lazar (2003) and Lancaster and Jun (2010) specifically considered EL. Most of the works were concerned with independent data analysis with few exceptions. Dunson et al. (2003) considered median regression for a latent variable model with multiple surrogate outcomes under Gaussian within-subject dependency structure. Yin (2009) used a quadratic likelihood-like function motivated by generalized estimating equations and proposed Bayesian generalized method of moments method.

The proposed method is similarly motivated: the aforementioned Bayesian nonparametric methods are complex in the computation and require an informative modeling of the error heteroscedasticity. This work does not require modeling of the error densities. EL is one choice that does not require modeling of the error densities. Any non-parametric likelihood such as exponentially tilted EL can be used instead. Using the MCMC sampler, the proposed method also does not require directly estimating the variance of the estimator for inference and is not subject to a known challenge of QR inference of estimating error densities at the quantile of interest. In this paper we provide large sample properties of the resulting quasi-posterior estimators and inference, being the first work clearly showing shrinkage of the random effects estimators toward the population average effect and the asymptotic normality of the population average effect estimator.

The remainder of this paper proceeds as follows. Section 2 formally defines the semi-parametric random effects quantile regression (REQR) estimator and provides their large sample properties. Section 3 describes MCMC methods. Section 4 provides empirical results including the analysis of two real data examples. Section 5 concludes. All the proofs are deferred to the Appendix.

2. MODELS AND METHODS

For expository purposes we assume random design points in this section. If the design points are non-stochastic, the results we obtain in this section hold under appropriate conditions on the design sequences.

2.1 Random Effects Quantile Regression (REQR) Model

Let $\{(y_{ij}, \mathbf{x}_{ij})\}_{j=1}^{m_i}$, $i = 1, \dots, n$, denote observations from n clusters with cluster size m_i . For each of the n clusters, we suppose that $\{(y_{ij}, \mathbf{x}_{ij})\}_{j=1}^{m_i}$ is a random sample from the following random effects quantile regression (REQR) model,

$$Q_{i,\tau}\{y|\mathbf{x}, \mathbf{b}_i(\tau)\} = \mathbf{x}^\top \mathbf{b}_i(\tau) \text{ for } \tau \in (0, 1), \quad (1)$$

where $\mathbf{x} \in \Xi_i \subseteq R^q$ is a random covariate that includes an intercept with a distribution function F_i^X , Ξ_i is the support of $\mathbf{x}$ in the i -th cluster, and $Q_{i,\tau}\{y|\mathbf{x}, \mathbf{b}_i(\tau)\}$ denotes the τ -th conditional quantile of the response $y \in R$ in the i -th cluster given $\mathbf{x}$ and $\mathbf{b}_i(\tau)$. We view the cluster as randomly chosen from a population, and thus the cluster-specific quantile coefficients $\mathbf{b}_i(\tau)$ are random variables. We further assume the below condition:

C.1 For $i = 1, \dots, n$, $\mathbf{b}_i(\tau)$ are q -dimensional random variables supported on $\mathcal{T}_i$ with a common mean $\boldsymbol{\beta}(\tau)$ and cluster-specific covariance matrix $\boldsymbol{\Sigma}_i(\tau)$ that are positive definite.

Similar to the usual QR analysis for independent data ($m_i = 1$), $\mathbf{b}_i(\tau)$ denotes the quantile specific effects of the covariates on the responses in the i -th cluster, while the common mean $\boldsymbol{\beta}(\tau)$ corresponds to the population-average effects of the covariates specific to the τ -th conditional quantile of the response. If $\boldsymbol{\Sigma}_i(\tau) = \boldsymbol{\Sigma}(\tau)$ for all clusters, $\boldsymbol{\Sigma}(\tau)$ represents the variability of the quantile specific covariate effects across the clusters. In this paper we consider the problem of estimating the model for a specific regression quantile with a given $\tau \in (0, 1)$. We therefore omit τ from notations whenever no confusion would arise.

In order to compare the model with the conditional mean model, we rewrite (1) as

$$y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \mathbf{x}_{ij}^\top \mathbf{b}_i^* + e_{ij}, \quad (2)$$

where $\mathbf{b}_i^*$ are mean zero random effects that correspond to deviations in the covariate effects of the τ -th conditional quantiles in the individual clusters from the population-average effect, and e_{ij} are random errors with zero τ -th conditional quantiles. Similar to the conditional mean regression, e_{ij} are independent of $\mathbf{b}_i^*$ given $\mathbf{x}_{ij}$. In the Network for the Improvement for Addiction Treatment (NIATx) example presented in Section 4.1 later, where $\tau = 0.9$ and $\mathbf{x} = (1, t, t^2)$ is considered with t denoting a random admission time counted from the beginning of the NIATx funding, $\mathbf{x}^\top \boldsymbol{\beta}$ corresponds to the mean of the 90th percentile wait time at the admission time t . Also $\mathbf{x}^\top \mathbf{b}_i^* < 0$ indicates a shorter wait time than the population average of the 90th percentile in the i -th addiction treatment center, while $\mathbf{x}^\top \mathbf{b}_i^* > 0$ indicates a longer wait time.

If $\tau = 0.5$ and e_{ij} have symmetric distributions with a finite mean, the median REQR is equal to the conditional mean: $\mathbf{x}_{ij}^\top \mathbf{b}_i(\tau) = E\{y_{ij} | \mathbf{x}_{ij}, \mathbf{b}_i(\tau)\}$. Moreover, the averaged conditional median is the conditional mean of a marginal model: $\mathbf{x}_{ij}^\top \boldsymbol{\beta}(\tau) = E(y_{ij} | \mathbf{x}_{ij})$ for $\tau = 0.5$. In this case, the proposed method can be seen as a robust alternative to the random effects conditional mean regression against outliers in the responses.

It is well known with the conditional mean regression that integrating over the random effects produces a marginal model where the regression parameters retain their meaning. Such a relationship does not hold with the REQR. Model (1) is not generally related to a marginal QR model, $Q_\tau(y|\mathbf{x}) = \mathbf{x}^\top \boldsymbol{\zeta}(\tau)$. Obviously, we cannot expect $\boldsymbol{\beta}(\tau) = \boldsymbol{\zeta}(\tau)$ unless $\Xi_i = \Xi$ and

$$E[F_i\{-\mathbf{x}^\top \mathbf{b}_i^*(\tau) | \mathbf{x}\} | \mathbf{x}] = \tau \text{ uniformly for all } \mathbf{x} \in \Xi \text{ and } 1 \leq i \leq n,$$

where $F_i\{\cdot | \mathbf{x}\}$ denote the conditional distribution of e_{ij} given $\mathbf{x}_{ij}$ in model (2) and the expectation is taken with respect to the distribution of $\mathbf{b}_i^*(\tau)$.

2.2 Estimation

Let $\Sigma_1, \dots, \Sigma_n$ be parameterized by a vector $\boldsymbol{\sigma}$. Define $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\sigma})$ with the true value denoted by $\boldsymbol{\theta}_0$ such that $\boldsymbol{\beta}(\boldsymbol{\theta}_0) = \boldsymbol{\beta}_0$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0) = \boldsymbol{\Sigma}_{i0}$. We denote the parameter space of $\boldsymbol{\theta}$ by Θ . We assume the following conditions for all $1 \leq i \leq n$:

- C.2** $\mathbf{b}_i$ is the unique solution to $E[\phi_\tau\{y - \mathbf{x}^\top \mathbf{b}_i\} \mathbf{x}] = 0$, where $\phi_\tau(t) = \tau - I_{[t < 0]}$, the derivative of the check function and the expectation is taken over the conditional distribution of y given $\mathbf{x}$ in the i -th cluster.
- C.3** $F_i(t|\mathbf{x})$ has the density $f_i(t|\mathbf{x})$ that is differentiable at 0 with $0 < f_i(t|\mathbf{x}) < \infty$ and $\sup_{\mathbf{x} \in \Xi_i, |t| < \varepsilon} |f'_i(t|\mathbf{x})| < \xi$ for $\xi > 0$ and $\varepsilon > 0$.
- C.4** $E(\mathbf{x}\mathbf{x}^\top) = \mathbf{D}_{i0}$ and $E\{f_i(0|\mathbf{x})\mathbf{x}\mathbf{x}^\top\} = \mathbf{D}_{i1}$ in the i -th cluster for some positive definite matrix $\mathbf{D}_{i0}$ and $\mathbf{D}_{i1}$, where the expectations are taken over $\mathbf{x}$ from F_i^X .
- C.5** Ξ_i is uniformly bounded in R^q .

Under conditions C.2–C.4 standard QR method applied to each cluster separately produces estimators of the random coefficients $\hat{\mathbf{b}}_i$ with asymptotic variances $m_i^{-1}\mathbf{Q}_i$ where $\mathbf{Q}_i = \mathbf{D}_{i1}^{-1}\mathbf{D}_{i0}\mathbf{D}_{i1}^{-1}$. We define

$$\mathbf{V}_i = \sum_i m_i^{-1} \mathbf{Q}_i. \quad (3)$$

Given $\mathbf{x}$, $\mathbf{V}_i$ is the asymptotic variance of the the per cluster standard QR analysis estimator $\mathbf{x}^\top \mathbf{b}_i$ about the population averaged regression quantile of interest: $E\{(\mathbf{x}^\top \hat{\mathbf{b}}_i - \mathbf{x}^\top \boldsymbol{\beta}^2 | \mathbf{x}) = \mathbf{x}^\top \mathbf{V}_i \mathbf{x}$. Condition C.5 is a sufficient condition and may be relaxed.

We first define a semiparametric likelihood-like criterion function for the random coefficients $\mathbf{b}_i$. Given τ and in the i -th cluster, $\hat{\mathbf{b}}_i$ is formally defined as a solution to the minimization problem

$$\min_{\mathbf{b} \in R^q} \sum_{j=1}^{m_i} \rho_\tau(y_{ij} - \mathbf{x}_{ij}^\top \mathbf{b}),$$

where $\rho_\tau(\cdot)$ is the check function. As well known, $\hat{\mathbf{b}}_i$ can be alternatively defined as a solution to estimating equations,

$$m_i^{-1/2} \sum_{j=1}^{m_i} \phi_\tau(y_{ij} - \mathbf{x}_{ij}^\top \mathbf{b}) \mathbf{x}_{ij} \approx 0.$$

This estimating equations formulation motivates the below empirical likelihood (EL) for the coefficient $\mathbf{b}_i$ in the i -th cluster:

$$L_{m_i}(\mathbf{b}) = \max \left\{ \prod_{j=1}^{m_i} p_j \sum_{j=1}^{m_i} p_j \phi_\tau(y_{ij} - \mathbf{x}_{ij}^\top \mathbf{b}) \mathbf{x}_{ij} = 0, \sum_{j=1}^{m_i} p_j = 1, 0 \leq p_j \leq 1 \right\},$$

where p_j denote weights for the j -th observation in the cluster respectively. Like in parametric case, EL method considers a parameter as a function of distributions, specifically those empirically defined by a set of weights $\{p_j\}_{j=1}^{m_i}$, and defines a likelihood for the parameter by profiling out $\{p_j\}_{j=1}^{m_i}$. We refer to Owen (2001) for a general introduction and

Qin and Lawless (1994) for its application with estimating equations. We note that the above EL formulation is also valid with fixed design points under appropriate conditions on the design sequence (see Owen (1991) for details), and accordingly the results of this paper hold for fixed design points.

We let $g_i(\cdot|\boldsymbol{\eta})$ denote the densities of $\mathbf{b}_i$ under Condition C.1, where $\boldsymbol{\eta}$ is a r -variate vector in Ψ parameter space and not necessarily equal to $\boldsymbol{\theta}$. For expository simplicity, we assume $\boldsymbol{\eta} = \boldsymbol{\theta}$ and $\Psi = \Theta$ here. We define a semiparametric likelihood-like criterion function for $\mathbf{b}_i$ by

$$\widetilde{L}_{m_i}(\mathbf{b}|\boldsymbol{\theta}) = L_{m_i}(\mathbf{b})g_i(\mathbf{b}|\boldsymbol{\theta}). \quad (4)$$

By the independence among the clusters a semiparametric likelihood-like criterion function for $\boldsymbol{\theta}$ is given by

$$\prod_{i=1}^n \int_{\mathbf{b}_i \in \Upsilon_i} \widetilde{L}_{m_i}(\mathbf{b}_i|\boldsymbol{\theta}) d\mathbf{b}_i.$$

Under some regularity conditions, the above integral admits a quadratic form of the first order expansion on the log scale centered at $\boldsymbol{\theta}_0$ in a $O(n^{-1/2})$ neighborhood of $\boldsymbol{\theta}_0$, while $\widetilde{L}_{m_i}(\mathbf{b}_i|\boldsymbol{\theta})$ is asymptotically equivalent to a multivariate normal density (see (A.6) and (A.7) in the Appendix). This motivates the following maximizers as natural estimators:

$$\bar{\mathbf{b}}_{m_i} = \arg\max_{\mathbf{b} \in \Upsilon_i} \widetilde{L}_{m_i}(\mathbf{b}_i|\bar{\boldsymbol{\theta}}_n), \quad \bar{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta} \in \Theta} \prod_{i=1}^n \int_{\mathbf{b}_i \in \Upsilon_i} \widetilde{L}_{m_i}(\mathbf{b}_i|\boldsymbol{\theta}) d\mathbf{b}_i.$$

A direct computation of the estimators is daunting. We instead use the Bayesian framework and MCMC samplers.

Let $\pi(\boldsymbol{\theta})$ denote a non-informative prior appropriately defined for $g_i(\cdot|\boldsymbol{\theta})$. Define

$$\omega_n(\boldsymbol{\theta}, \mathbf{b}_1, \dots, \mathbf{b}_n) = \prod_{i=1}^n \widetilde{L}_{m_i}(\mathbf{b}_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

We consider a quasi-posterior defined by $\omega_n(\boldsymbol{\theta}, \mathbf{b}_1, \dots, \mathbf{b}_n)$:

$$\tilde{\omega}_n(\boldsymbol{\theta}, \mathbf{b}_1, \dots, \mathbf{b}_n) = \frac{\omega_n(\boldsymbol{\theta}, \mathbf{b}_1, \dots, \mathbf{b}_n)}{\int_{\boldsymbol{\theta} \in \Theta} \int_{\mathbf{b}_1 \in \Upsilon_1} \dots \int_{\mathbf{b}_n \in \Upsilon_n} \omega_n(\boldsymbol{\theta}, \mathbf{b}_1, \dots, \mathbf{b}_n) d\mathbf{b}_n \dots d\mathbf{b}_1 d\boldsymbol{\theta}}. \quad (5)$$

Improper priors may not guarantee a proper posterior and Yang and He (2011) specifically advise avoiding flat priors with their Bayesian EL approach. With the proposed REQR method, however, improper priors will guarantee a proper posterior, if they produce a proper posterior in relation to $g_i(\mathbf{b}|\boldsymbol{\theta})$ when applied to $\{\mathbf{b}_i\}_{i=1}^n$ as if they were observed. From the definition of EL, we note that $\limsup_{m_i \rightarrow \infty} \sup_{\mathbf{b} \in \Upsilon_i} L_{m_i}(\mathbf{b}) \leq 1$ for all $1 \leq i \leq n$ and every n . We have

$$\int_{\theta \in \Theta} \int_{\mathbf{b}_1 \in \Upsilon_1} \cdots \int_{\mathbf{b}_n \in \Upsilon_n} \omega_n(\theta, \mathbf{b}_1, \dots, \mathbf{b}_n) d\mathbf{b}_n \cdots d\mathbf{b}_1 d\theta \leq \int_{\theta \in \Theta} \int_{\mathbf{b}_1 \in \Upsilon_1} \cdots \int_{\mathbf{b}_n \in \Upsilon_n} \prod_{i=1}^n g_i(\mathbf{b}_i | \theta) \pi(\theta) d\mathbf{b}_n \cdots d\mathbf{b}_1 d\theta,$$

where the second integral can be finite with a certain choice of improper $\pi(\theta)$ specific to $g_i(\mathbf{b}_i | \theta)$. For example, when $g_i(\mathbf{b}_i | \theta) = \varphi(\mathbf{b}_i | \boldsymbol{\beta}, \boldsymbol{\Sigma})$ for some normal density $\varphi(\cdot)$, a joint flat prior is the improper prior $\pi(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(q+1)/2}$, which produces a proper posterior:

$$\sum \mathbf{b}_1, \dots, \mathbf{b}_n \sim IW_{n-1}(\mathbf{S}_n), \quad \beta | \sum \mathbf{b}_1, \dots, \mathbf{b}_n \sim N_{(q)} \left(n^{-1} \sum_{i=1}^n \mathbf{b}_i, n^{-1} \sum \right), \quad (6)$$

where $\mathbf{S}_n = \sum_{i=1}^n (\mathbf{b}_i - n^{-1} \sum_{i=1}^n \mathbf{b}_i) (\mathbf{b}_i - n^{-1} \sum_{i=1}^n \mathbf{b}_i)^\top$ and $IW_{n-1}(\cdot)$ denotes the inverse-Wishart distribution with $(n-1)$ degrees of freedom. If such choice of improper priors is not available, we recommend vague priors.

We let $\tilde{\omega}_{ni}(\mathbf{b})$ and $\tilde{\omega}_n(\theta)$ denote the marginal densities appropriately defined corresponding to $\mathbf{b}_i$ and θ . The quasi-posterior means are then respectively defined as

$$\tilde{\mathbf{b}}_{mi} = \int_{\mathbf{b} \in \Upsilon} \mathbf{b} \tilde{\omega}_{ni}(\mathbf{b}) d\mathbf{b}, \quad \tilde{\theta}_n = \int_{\theta \in \Theta} \theta \tilde{\omega}_n(\theta) d\theta. \quad (7)$$

We consider the following specific conditions in addition to Condition C.1.

C.1.(i) For $1 \leq i \leq n$, $g_i(\cdot | \boldsymbol{\eta}) = g(\cdot | \boldsymbol{\eta})$ with $\boldsymbol{\eta}_0$ denoting the true value.

- a. $\boldsymbol{\theta} = \mathbf{h}(\boldsymbol{\eta})$ for a vector of smooth functions $\mathbf{h}(\cdot)$ such that $\mathbf{H}(\boldsymbol{\eta}) = \partial \mathbf{h} / \partial \boldsymbol{\eta}$ is non-singular and has continuous entries in a neighborhood of $\boldsymbol{\eta}_0$ and $\boldsymbol{\theta}_0 = \mathbf{h}(\boldsymbol{\eta}_0)$.
- b. $0 < g(\mathbf{b} | \boldsymbol{\eta}) < \infty$ uniformly in $\mathbf{b} \in \Upsilon$ for any $\boldsymbol{\eta} \in \Psi$ and $g(\mathbf{b} | \boldsymbol{\eta})$ is two times continuously differentiable. Also $E \{ -\partial^2 \log g / (\partial \boldsymbol{\eta} (\partial \boldsymbol{\eta})^\top) |_{\boldsymbol{\eta}=\boldsymbol{\eta}_0} \}$ is positive definite. For a unit-vector $\mathbf{u}$, $|E\{\mathbf{u}^\top (\partial^2 \log g / (\partial \boldsymbol{\eta} (\partial \boldsymbol{\eta})^\top) \mathbf{u})\}| < \infty$ for all $\boldsymbol{\eta} \in \Psi$, and $|\mathbf{u}^\top E\{[\partial^2 \log g / (\partial \boldsymbol{\eta} (\partial \boldsymbol{\eta})^\top) |_{\boldsymbol{\eta}=\boldsymbol{\eta}_0} - \{\partial^2 \log g / (\partial \boldsymbol{\eta} (\partial \boldsymbol{\eta})^\top) |_{\boldsymbol{\eta}=(\boldsymbol{\eta}_0+\boldsymbol{\varepsilon})}\}] \mathbf{u}\}| = O(\|\boldsymbol{\varepsilon}\|)$, as $\|\boldsymbol{\varepsilon}\| \rightarrow 0$.

C.1.(ii) $\mathbf{b}_i$ have normal densities with not necessarily identical $\boldsymbol{\Sigma}_i$. The heterogeneity among $\boldsymbol{\Sigma}_i$ is limited such that $\boldsymbol{\theta}$ is a finite vector.

The requirements of condition C.1.(i) are comparable to those required for a general form of the Bayesian equivalent of the central limit theorem to hold for the posterior distribution of $\boldsymbol{\eta}$ when standard Bayesian method is applied to $\{\mathbf{b}_i\}_{i=1}^n$ if $\mathbf{b}_i$ were observed (see Chapter 5.3 of Bernardo and Smith (2003) for example). They are general and admit skewed multivariate normal or multivariate student-t distributions for $\mathbf{b}_i$. The limited heterogeneity of the normal densities in condition C.1.(ii) is a standard assumption of the Gaussian conditional mean regression analysis. It essentially enables $\boldsymbol{\Sigma}_i$ to be estimated at a parametric rate.

We first present asymptotic results with the normal densities. We define $\mathbf{V}_{i0} = \sum_{i0} + m_i^{-1} \mathbf{Q}_i$ following the definition of $\mathbf{V}_i$ in (3).

Theorem 1—Assume condition C.1, C.1.(ii) and C.2–C.5.

- i. As $m_i \rightarrow \infty$, $\tilde{\mathbf{b}}_{m_i} = \{\mathbf{I} - (m_i^{-1} \mathbf{Q}_i) \mathbf{V}_{i0}^{-1}\} \hat{\mathbf{b}}_i + (m_i^{-1} \mathbf{Q}_i) \mathbf{V}_{i0}^{-1} \tilde{\beta}_n + o_p(m_i^{-1/2})$ for all $1 \leq i \leq n$. Moreover, $\text{Var}(\tilde{\mathbf{b}}_{m_i}) = \{\mathbf{I} - (m_i^{-1} \mathbf{Q}_i) \mathbf{V}_{i0}^{-1}\} (m_i^{-1} \mathbf{Q}_i) + o_p(m_i^{-1})$.
- ii. If $m_i \rightarrow \infty$ for all $1 \leq i \leq n$ and $n \rightarrow \infty$, then $\tilde{\beta}_n = n^{-1} \sum_{i=1}^n (\bar{\mathbf{V}}_n \mathbf{V}_{i0}^{-1}) \hat{\mathbf{b}}_i + o_p(n^{-1/2})$, where $\bar{\mathbf{V}}_n = \{n^{-1} \sum_{i=1}^n \mathbf{V}_{i0}^{-1}\}^{-1}$. Also $n^{1/2}(\tilde{\beta}_n - \tilde{\beta}_0) \sim AN_{(q)}(0, \mathbf{V}_n)$.
- iii. Suppose $\Sigma_i = \Sigma$ and $\tilde{\Sigma}_n$ is accordingly defined by the posterior mean $\tilde{\sigma}_n$. Consider matrix operator $\text{vec}(\cdot)$ that takes a matrix and forms a large column vector by stacking the columns, and its generalization $\text{vech}(\cdot)$ for a symmetric matrix, respectively described in Henderson and Searle (1979). If $m_i \rightarrow \infty$ for all $1 \leq i \leq n$ and $n \rightarrow \infty$, then

$$n^{1/2} \{\text{vech}(\tilde{\Sigma}_n) - \text{vech}(\Sigma_0)\} \sim AN_{(q(q+1)/2)} \left(\mathbf{0}, 2 \left\{ \mathbf{K}^\top \left(n^{-1} \sum_{i=1}^n \mathbf{V}_{i0}^{-1} \otimes \mathbf{V}_{i0}^{-1} \right) \mathbf{K} \right\}^{-1} \right),$$

where $\mathbf{K}$ is a unique matrix given by the relationship $\text{vec}(\mathbf{M}) = \mathbf{K} \text{vech}(\mathbf{M})$ for any symmetric matrix $\mathbf{M}$ of the same dimension as Σ and $\otimes$ represents the Kronecker product.

The first result of the theorem shows shrinkage in the REQR estimators that is comparable to the conditional mean regression: given $\mathbf{x}$, the posterior means $\mathbf{x}^\top \tilde{\mathbf{b}}_{m_i}$ are weighted averages of the population averaged profile $\mathbf{x}^\top \tilde{\beta}_n$ and the within cluster QR estimates $\mathbf{x}^\top \hat{\mathbf{b}}_i$ with the weights defined by ratios between the asymptotic within cluster variances $m_i^{-1} \mathbf{Q}_i$ and $\mathbf{V}_{i0}$. Much weight will be given to the overall average profile if the within cluster variability of the quantile is large in comparison to the between-cluster variability, whereas much weight will be given to the cluster level estimator if the opposite is true. Also $\text{Var}(l^\top \tilde{\mathbf{b}}_{m_i}) \leq \text{Var}(l^\top \hat{\mathbf{b}}_i)$ asymptotically for any linear combination l of the random effects.

From (ii) we note that $\tilde{\beta}_n$ is a weighted average of the per cluster estimators $\hat{\mathbf{b}}_i$ with the weights determined by the relative magnitude of $\mathbf{V}_{i0}^{-1}$. If $\Sigma_{i0} = \Sigma_0$, then

$$\bar{\mathbf{V}}_n = \sum_0^{1/2} \left\{ \mathbf{I} - n^{-1} \sum_{i=1}^n (I + m_i \sum_0^{1/2} \mathbf{Q}_i^{-1} \sum_0^{1/2})^{-1} \right\}^{-1} \sum_0^{1/2}. \text{ As } \mathbf{Q}_i \text{ are positive definite,}$$

$$\text{Var}(l^\top \tilde{\beta}_n) \leq \text{Var}\{l^\top (n^{-1} \sum_{i=1}^n \hat{\mathbf{b}}_i)\} \text{ asymptotically for any linear combination } l.$$

In practice we have Markov chains $\mathbf{S}(\mathbf{b}_i) = (\mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)}, \dots, \mathbf{b}_i^{(B)})$, $i = 1, \dots, n$, and $\mathbf{S}(\theta) = (\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(B)})$ for some large B . We estimate $\mathbf{b}_{m_i}$, $\tilde{\beta}_n$, $\text{Var}(\mathbf{b}_{m_i})$, $\text{Var}(\tilde{\beta}_n)$, and $\text{Var}(\text{vech}(\tilde{\Sigma}_n))$ by the mean and the variance-covariance matrix of the corresponding MCMC sequences. For a given $\alpha \in (0, 1)$ we can construct $100(1 - \alpha)\%$ confidence intervals based on the normal approximation results given in Theorem 1. We denote this confidence intervals by $[c_{1,n}\{\mathbf{S}(\cdot), \alpha/2\}, c_{1,n}\{\mathbf{S}(\cdot), 1 - \alpha/2\}]$. Alternatively we can construct confidence intervals using the sample percentiles of the MCMC sequence $\mathbf{S}(\cdot)$. We denote them by $[c_{2,n}\{\mathbf{S}(\cdot), \alpha/2\}, c_{2,n}\{\mathbf{S}(\cdot), 1 - \alpha/2\}]$. The below theorem shows that both intervals are asymptotically valid.

Theorem 2—Assume condition C.1, C.1(ii) and C.2–C.5. For any $\alpha \in (0, 1)$,

- i. $\lim_{m_i \rightarrow \infty} P[c_{k,n}\{\mathbf{S}(\mathbf{b}_i), \alpha/2\} \leq \mathbf{b}_i \leq c_{k,n}\{\mathbf{S}(\mathbf{b}_i), 1 - \alpha/2\}] = 1 - \alpha$ for $k = 1, 2$ and $1 \leq i \leq n$.

- ii. $\lim_{n \rightarrow \infty} P[c_{k,n}\{\mathbf{S}(\boldsymbol{\theta}), \alpha/2\} \leq \boldsymbol{\theta}_0 \leq c_{k,n}\{\mathbf{S}(\boldsymbol{\theta}), 1 - \alpha/2\}] = 1 - \alpha$ for $k = 1, 2$.

The results similarly hold for the inference of $v(\mathbf{b}_i)$ or $v(\boldsymbol{\theta}_0)$ for some given continuously differentiable function $v: \Theta \rightarrow R$ or $v: \mathbb{T}_i \rightarrow R$. A known challenge of QR inference is that a direct estimation of the variance requires estimating the error densities at the quantile of interest. The proposed method is not subject to this requirement.

Here each $\widetilde{L}_{m_i}(\mathbf{b}_i|\theta)$ can be seen as a posterior density with $g_i(\mathbf{b}_i|\theta)$ an informative prior, and its asymptotic normality is shown by Lazar (2003) and Yang and He (2011). As $\widetilde{L}_{m_i}(\mathbf{b}_i|\theta)$ does not contain any population level parameters for the error distributions, completely allowing the errors to be heteroscedastic, and is only known to be consistent asymptotically, m_i are required to grow along with n for the asymptotic results of $\boldsymbol{\theta}$. With the normal densities assumed for $\mathbf{b}_i$, nevertheless, m_i are only needed to grow, independently of each other and independent of n . With the non-normal densities, however, we require a rather stricter condition $(\min_{1 \leq i \leq n} m_i)n^{-1/2} \rightarrow \infty$ as a sufficient condition as shown below.

Under the non-normal density condition of C.1.(i), we define $\mathbf{V}_{i0}$ with $\boldsymbol{\Sigma}_{i0} = \boldsymbol{\Lambda}(\boldsymbol{\eta}_0)$ where $\boldsymbol{\Lambda}(\boldsymbol{\eta})$ is given in Lemma A.3 in the Appendix.

Theorem 3—Assume Conditions C.1, C.1.(i) and C.2–C.5.

- i. As $m_i \rightarrow \infty$, $\tilde{\mathbf{b}}_{m_i} = \{\mathbf{I} - (m_i^{-1} \mathbf{Q}_i) \mathbf{V}_{i0}^{-1}\} \widehat{\mathbf{b}}_i + (m_i^{-1} \mathbf{Q}_i) \mathbf{V}_{i0}^{-1} \tilde{\beta}_n + o_p(m_i^{-1/2})$ for all $1 \leq i \leq n$.
Moreover, $\text{Var}(\tilde{\mathbf{b}}_{m_i}) = \{\mathbf{I} - (m_i^{-1} \mathbf{Q}_i) \mathbf{V}_{i0}^{-1}\} (m_i^{-1} \mathbf{Q}_i) + o_p(m_i^{-1})$.
- ii. If $n \rightarrow \infty$ along with m_i , $1 \leq i \leq n$, such that $(\min_{1 \leq i \leq n} m_i)n^{-1/2} \rightarrow \infty$, then
 $n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \sim AN\left(0, \left[\mathbf{H}(\boldsymbol{\eta}_0) \mathbf{J}_{\boldsymbol{\eta}_0} \boldsymbol{\Omega}_{\boldsymbol{\eta}_0}^{-1} \mathbf{J}_{\boldsymbol{\eta}_0}^{\top} \{\mathbf{H}(\boldsymbol{\eta}_0)\}^{\top}\right]^{-1}\right)$, where $\mathbf{H}(\cdot)$ is the Jacobian given in the Condition C.1.(i), and $\mathbf{J}_{\boldsymbol{\eta}_0}$ and $\boldsymbol{\Omega}_{\boldsymbol{\eta}_0}$ are quantities related to the hessian and the score of $\log \omega_n(\boldsymbol{\eta})$ defined in (A.6) in the Appendix.

Inference results corresponding to Theorem 2 also holds accordingly.

3. MARKOV CHAIN MONTE CARLO (MCMC)

We propose a simple combination of Metropolis-Hastings (M-H) and Gibbs algorithm. As $\mathbf{b}_i$ are independent given $\boldsymbol{\theta}$ and the posterior is always proportional to the joint, we have

$$\tilde{\omega}_n(\mathbf{b}_i|\theta) \propto \omega_n(\theta, \mathbf{b}_1, \dots, \mathbf{b}_n) \propto \widetilde{L}_{m_i}(\mathbf{b}_i|\theta) = L_{m_i}(\mathbf{b}_i) g_i(\mathbf{b}_i|\theta) \text{ for all } 1 \leq i \leq n,$$

where $\propto$ denotes equivalence up to some constants. Many algorithms are available for the computation of $L_{m_i}(\mathbf{b}_i)$, for example, R package ‘emplik’. We use a M-H algorithm with a pre-specified normal proposal distribution to sample each $\mathbf{b}_i$. On the other hand,

$$\tilde{\omega}_n(\theta|\mathbf{b}_1, \dots, \mathbf{b}_n) = \frac{\tilde{\omega}(\theta, \mathbf{b}_1, \dots, \mathbf{b}_n)}{\int_{\theta \in \Theta} \tilde{\omega}(\theta, \mathbf{b}_1, \dots, \mathbf{b}_n) d\theta} \propto \prod_{i=1}^n g_i(\mathbf{b}_i|\theta) \pi(\theta),$$

where the right hand side may have a readily available closed form as $g_i(\mathbf{b}_i|\theta)$ are parametrically specified. A Gibbs sampling or M-H algorithm is used to sample $\boldsymbol{\theta}$. Specifically when $\mathbf{b}_i$ are assumed having a common Gaussian distribution as in the empirical studies in Section 4, $\boldsymbol{\theta} = (\boldsymbol{\Sigma}, \boldsymbol{\mu})$ can be sampled by a Gibbs sampling algorithms using an

inverse-Wishart and a normal density as shown in (6). For expository simplicity, we assume $\eta = \theta$ above. If $\eta \neq \theta$, η will be sampled instead of θ and θ is given by $\mathbf{h}(\eta)$ under Condition C.1.(i). As mentioned previously, we avoid modeling the error densities entirely, and the MCMC computation is much simpler.

4. NUMERICAL STUDIES

All computations were conducted in R (R Development Core Team, 2006) with the empirical likelihood computed using `el.cen.EM2` function in R package ‘`emplik`’. Proposal distributions for the Metropolis-Hastings algorithm were chosen to have acceptance rates in the range of 0.1 ~ 0.4.

4.1 Real Data Examples

Real data examples were analyzed with 25,000 MCMC samples, the first 5,000 discarded as burn-in, and the standard QR solutions as initial values. The random effects were assumed Gaussian. The convergence of the Markove chains was checked and found satisfactory by the methods recommended by Cowles and Carlin (1996). For example, based on the Z-score by Geweke (1992), we found that all Z-scores were not significant. Specifically for the Example 1 NIATx data, Z-scores took values of 1.217, -0.074, and -1.747 for $\beta = (\beta_0, \beta_1, \beta_3)$ and 1.699, 0.243, -1.514, 0.261, 1.520, and 0.544 for $\Sigma = (\Sigma_{11}, \Sigma_{12}, \Sigma_{13}, \Sigma_{22}, \Sigma_{23}, \Sigma_{33})$ respectively. Z-scores for the random effects ranged from -1.897 to 1.723. We provide the trace plots in Figure 1.

Example 1—NIATx We analyze data from the Network for the Improvement for Addiction Treatment (NIATx) using a second degree polynomial REQR. The NIATx is a learning community of alcohol and other drug treatment centers. It uses process improvement strategies to treatment services and facilitates organizational change in the delivery of addiction treatment services (Capoccia et al. (2007)). A cross-center evaluation monitored impacts on days between first contact and first treatment. In this paper we analyze data on the delay time from twenty-six programs that provided at least ten months of data. The majority of the centers (22/26) have samples larger than 150 with eighteen of them greater than 200. This allows modeling of high quantiles as shown below. We refer to McCarty et al. (2007) and Hoffman et al. (2008) for more information on the study design, methods and outcomes.

Centers that successfully used process improvement reduced delay times and the reduction was more apparent the longer they implemented process improvements. The decreasing trend in the delay time, however, is not adequately summarized by the conditional mean. A small or zero improvement is expected in the lower tail of the distribution where delays were previously short. In the upper tail where delays were long, however, a more pronounced improvement is expected. The upper half of the data above the median are particularly interesting as the process improvement strategies were anticipated to reduce the spread of the data towards the median in the upper half. This justifies REQR analysis of the data. The distribution of delay times is right skewed (see examples in Figure 2) and analyzed on the log scale. We include admission time of each patient (in month since the beginning of NIATx participation) as a covariate and consider a change in the delay time as a second degree polynomial function of the admission time. For comparison we analyze the data for changes in the conditional mean delay time using restricted maximum likelihood (REML) method. We also conduct QR analysis separately for each center using the standard independent data QR method.

Figure 2 shows the regression results for four centers. Increasing heterogeneity is observed among the centers at higher quantiles. The increasing between center heterogeneity is more

clearly depicted in Figure 3. If the success of the process improvement strategies is reflected in decreasing delay time, this suggests that successful centers differ from unsuccessful ones more in the upper quantiles than in the medians. Center 24 is an example where the conditional mean analysis found no significant change, while the proposed REQR method found significant trends locally in the high quantiles. Center 12 exhibits an ideal trend anticipated with the process improvement strategies. The standard independent data QR method is known to sometimes produce crossing quantile fits in a small sample and Center 73 is one such case. Quantiles estimated by the proposed REQR, however, are not crossed. The REQR analysis pools information across the centers and the borrowed information offsets the small sample size.

Figure 4 presents point and interval estimates of the random coefficients of the linear term in the second degree polynomial 0.9-th quantile curves by the independent data QR and proposed RERQ method. Horizontal lines are respectively the population average effect estimate by the REQR method and a simple average of the per center estimates by the independent data QR method. The point estimates by the RERQ method are shrunk toward the population average effect estimate, and the interval estimates by the proposed RERQ method tend to be shorter than those by the standard independent QR method. This supports the asymptotic results of Theorem 1-(i).

Example 2—Swallow Study In this section we analyze apnea duration data from swallow study previously analyzed in Reich et al. (2010) and reference therein. Apnea duration is the period of nasal airflow cessation during swallowing. Longer durations often indicate age-related function changes for seniors. The duration was measured repeatedly under different experimental conditions defined by feeding condition (examiner- versus self-fed), viscosity (liquid or pudding), and volume (5ml or 10ml). Similar to Reich et al. (2010), we include indicators of the three experimental conditions and their two-way interactions as covariates and analyze a total of 1286 observations measured on twenty-three elderly women of age seventy-five or older, each with fifty-five or fifty-six observations.

The distribution of apnea duration times remains right-skewed after a log transformation and the spread of the data also varies widely across the subjects (see examples in Figure 5). Figure 6 shows that the data distributions under different experimental conditions differ more in the right tail than in the middle. These justify the proposed REQR analysis. We analyze the data for the conditional median and 0.9-th quantile. For comparison we conduct standard QR analysis that assumes all observations are independent.

Table 1 provides a summary. The REQR results agree well with the boxplot presentation of the data in Figure 6. The significant viscosity and viscosity by feeding condition interaction estimates suggest that swallowing liquid took significantly longer time, particularly when the subjects were fed by examiners. This is indicated by the higher median and 0.9-th sample quantile lines in the white boxes than in the shaded boxes across the experimental conditions, while the differences much bigger in the comparison of the wide white boxes with the wide shaded boxes. The non-significant volume and significant viscosity by volume interaction estimates suggest that swallowing the larger quantity did not necessarily take significantly more time, while the difference in the duration time by the volume was significantly larger when liquid was swallowed compared to pudding. This is implied by the greater height differences in the median and 0.9-th sample quantile lines by volume among the white boxes than the shaded boxes in Figure 6. The estimated effects are much larger for the 0.9-th conditional quantile.

Compared to the independent data QR analysis, the REQR provides similar results, mostly agreeing on the significance of the effects implied by the interval estimates. However, the

interval estimates tend to be much shorter with the REQR. The shorter intervals are expected as the independent data QR method ignores the within subject correlation in the data. The REQR also finds one more significant effect in the 0.9-th conditional quantile analysis.

We now compare our results with those of Reich et al. (2010). The results are notably different concerning the effects of feeding condition and viscosity by feeding interaction in the 0.9-th conditional quantile analysis. Contrary to our results, Reich et al. (2010) found the feeding condition effect significant, while the viscosity by feeding interaction effect non-significant. The two results are not directly comparable as Reich et al. (2010) considered the population quantiles using the below marginal model:

$$y_{ij} = \mathbf{x}_{ij}^T \eta(\tau) + \mathbf{x}_{ij}^T \gamma(\alpha_i(\tau) + e_{ij}),$$

where $\alpha_i(\tau)$ are random subject effects, accounting for the within subject dependency, and e_{ij} are independent within subject errors. As discussed in Section 2.1, our conditional model and the above marginal model are not necessarily related and cannot be compared. Nevertheless we note that our results agree with the boxplot representation of the data better.

4.2 Simulations

We use the below model and study the performance of the proposed REQR method by Monte-Carlo simulation:

$$y_{ij} = b_{0i} + b_{1i} \times z_{ij} + e_{ij}, i = 1, \dots, n; j = 1, \dots, m,$$

where b_{0i} and b_{1i} are univariate random variables, respectively sampled from an independent standard normal distribution $N_{(1)}(\beta_0, \sum_0^2)$ and $N_{(1)}(\beta_1, \sum_1^2)$, z_{ij} are sampled from a uniform distribution, $U(0, 12^{1/2})$, and e_{ij} are independent of b_{0i} and b_{1i} , sampled from one of the following six distributions: $N_{(1)}(0, 1)$, log normal distribution ($LN(0, 1)$), t distribution with the degree of freedom 2 (t_2), and their heteroscedastic versions.

4,000 MCMC samples were drawn per replicated data and the first 1,000 were discarded as burn-in. We randomly sampled values from $N_{(1)}(0, 1)$ and used them as initial values. In each simulation setting we examined the convergence of the Markove chains for the first 10 replicated datasets and found the convergence satisfactory similarly by the methods recommended by Cowles and Carlin (1996). We provide examples of trace plots in Figure 7.

The results for the medians ($\tau = 0.5$) are summarized and compared with the results of the restricted maximum likelihood (REML) method in Table 2 and 3. When the ho-moscedastic Gaussian error assumption of the REML is met, the REQR performed slightly worse in a small sample ($m = 10$) yet comparably in a moderately large sample ($m = 20$). The REQR and REML results are not directly comparable about the intercept when the error distribution is log normal, as they measure different quantities. With this exception, the proposed method outperformed the REML, particularly when inference is concerned with the variances. In a small sample ($m = 10$) the REML reported 0 for the variances of the random effects particularly in an alarmingly large number of cases for the heteroscedastic errors. With the 0 random effects variance estimates, the error variance estimate is inflated. This leads to a poor coverage for the variance parameters and an over-coverage and extended confidence intervals for the mean parameters β . The REQR performed overall decently in a small sample ($m = 10$) with homogeneous errors and reliably across all distributions in a moderate sample ($m = 20$) across all the error distributions.

Table 4 shows that the proposed method performed reasonably well for the 0.9-th quantile in a small sample with homoscedastic errors yet only for the slope parameters with heteroscedastic errors. Increase in the sample size did not improve the performance much for the intercept parameter. The non-normal distributions under consideration are extreme examples and represent challenging cases under the heteroscedastic setup. Moreover the simulation setups are not favorable with small between cluster variabilities relative to the within cluster variabilities. The high quantile under consideration compounded the difficulty. The performance improved when the between cluster variability increased. For example, when $b_{0i} \sim N_{(1)}(0, 4)$ and $b_{1i} \sim N_{(1)}(0, 4)$, we obtained comparable coverage results under the heteroscedastic settings as to the homoscedastic setting results reported in Table 4.

5. CONCLUSION

In this paper we extend QR method to a random effects analysis of clustered data, while fully preserving the features of the standard independent data QR method and admitting non-Gaussian random effects. The parameterization of the proposed REQR model is intuitive. The mean of the cluster specific random coefficients correspond to the population-average effects of the explanatory variables on the conditional quantile of interest. The discrepancy between the random coefficients and their mean corresponds to the deviation in the explanatory variable effects in the individual clusters from the population average. With symmetric error distributions with a finite mean, the median REQR coincides with the conditional mean and can be seen as a robust alternative that performs comparably when the homoscedastic Gaussian assumptions are met and outperforms in other cases. The proposed approach also avoids modeling the error distributions entirely and is computationally simpler than existing nonparametric Bayesian methods.

Acknowledgments

The authors thank Drs. Dongseok Choi and Dennis McCarty in the Department of Public Health and Preventive Medicine at the Oregon Health & Science University for sharing the Network for the Improvement for Addiction Treatment (NIATx) data. The authors thank the editor, the associate editor and two reviewers for their insightful comments. This research is supported in part by the National Institutes of Health (1R03CA133944-01). The collection of the NIATx data is supported by awards from the Robert Wood Johnson Foundation, the Center for Substance Abuse Treatment, and the National Institute on Drug Abuse (R01 DA018282; R01 DA020832).

References

- Bernardo, JM.; Smith, AFM. Bayesian Theory. John Wiley & Sons, Ltd; 2003.
- Capoccia VA, Cotter F, Gustafson DH, Cassidy E, Ford J, Madden L, Owens BH, Farnum SO, McCarty D, Molfenter T. Making ‘Stone Soup’: Improvements in Clinic Access and Retention in Addiction Treatment. Joint Commission Journal on Quality and Patient Safety. 2007; 33:95–103. [PubMed: 17370920]
- Chernozhukov V, Hong H. An MCMC Approach to Classical Estimation. Journal of Econometrics. 2003; 115:293–346.
- Cowles MK, Carlin BP. Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. Journal of American Statistical Association. 1996; 91:883–904.
- Dunson DB, Taylor JA. Approximate Bayesian Inference for Quantiles. Journal of Nonparametric Statistics. 2005; 17:385–400.
- Dunson DB, Watson M, Taylor JA. Bayesian Latent Variable Models for Median Regression on Multiple Outcomes. Biometrics. 2003; 59:296–304. [PubMed: 12926714]
- Geraci M, Bottai M. Quantile Regression for Longitudinal Data Using the Asymmetric Laplace Distribution. Biostatistics. 2007; 8:140–154. [PubMed: 16636139]

- Geweke, J. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In: Bernardo, JM.; Berger, J.; Dawid, AP.; Smith, AFM., editors. Bayesian Statistics. Vol. 4. Oxford: Oxford University Press; 1992. p. 169-193.
- Henderson HV, Searle SR. Vec and Vech Operators for Matrices, With Some Uses in Jacobinas and Multivariate Statistics. The Canadian Journal of Statistics. 1979; 7:65–81.
- Hanson T, Johnson WO. Modeling Regression Error with a Mixture of Polya trees. Journal of the American Statistical Association. 2002; 97:1020–1033.
- Hoffman KA, Ford JH, Choi D, Gustafson DH, McCarty D. Replication and Sustainability of Improved Access and Retention Within the Network for the Improvement of Addiction Treatment. Drug and Alcohol Dependence. 2008; 98:63–69. [PubMed: 18565693]
- Jung S. Quasi-likelihood for Median Regression Models. Journal of the American Statistical Association. 1996; 91:251–257.
- Koenker R. Quantile Regression for Longitudinal Data. Journal of Multivariate Analysis. 2004; 91:74–89.
- Koenker, R. Quantile Regression. Cambridge University Press; 2005.
- Kottas A, Gelfand AE. Bayesian Semiparametric Median Regression Modeling. Journal of the American Statistical Association. 2001; 96:1458–1468.
- Kottas A, Krnjajic M. Bayesian Nonparametric Modeling in Quantile Regression. Scandinavian Journal of Statistics. 2009; 36:297–319.
- Lavine M. On an Approximate Likelihood for Quantiles. Biometrika. 1995; 82:220–222.
- Lazar NA. Bayesian Empirical Likelihood. Biometrika. 2003; 90:319–326.
- Lancaster T, Jun SJ. Bayesian quantile regression methods. Journal of Applied Econometrics. 2010; 25:287–307.
- McCarty D, Gustafson DH, Wisdom JP, Ford J, Choi D, Molfenter T, Capoccia V, Cotter F. The Network for the Improvement of Addiction Treatment (NIATx): Enhancing access and retention. Drug and Alcohol Dependence. 2007; 88:138–145. [PubMed: 17129680]
- McCulloch CE. Symmetric Matrix Derivatives with Applications. Journal of the American Statistical Association. 1982; 77:679–682.
- Owen A. Empirical Likelihood for Linear Models. The Annals of Statistics. 1991; 19:90–120.
- Owen, A. Empirical Likelihood. Chapman & Hall/CRC Press; 2001.
- Qin J, Lawless J. Empirical Likelihood and General Estimating Equations. The Annals of Statistics. 1994; 22:300–325.
- Reich BJ, Bondell HD, Wang HJ. Flexible Bayesian Quantile Regression for Independent and Clustered Data. Biostatistics. 2010; 11:337–352. [PubMed: 19948746]
- Ruppert D, Carroll RJ. Trimmed Least Squares Estimation in the Linear Model. Journal of the American Statistical Association. 1980; 75:828–838.
- Schennach SM. Bayesian Exponentially Tilted Empirical Likelihood. Biometrika. 2005; 92:31–46.
- Tian L, Liu JS, Wei LJ. Implementation of Estimating Function-Based Inference Procedures With Markov Chain Monte Carlo Samplers. Journal of the American Statistical Association. 2007; 102:881–888.
- Yin G. Bayesian Generalized Method of Moments. Bayesian Analysis. 2009; 4:191–208.
- Yu K, Moyeed RA. Bayesian Quantile Regression. Statistics and Probability Letters. 2001; 54:437–447.
- Yang Y, He X. Bayesian Empirical Likelihood for Quantile Regression. The Annals of Statistics. 2011 under review.

APPENDIX: PROOFS

We will omit τ from notations below wherever the omission does not cause confusion. We let $\mathbf{z}(\mathbf{x}_{ij}, y_{ij}, \mathbf{b}) = \mathbf{x}_{ij} \phi_{\tau}(y_{ij} - \mathbf{x}_{ij}^T \mathbf{b})$, $l_{m_i}(\mathbf{b}) = \log L_{m_i}(\mathbf{b})$ and $\widetilde{l}_{m_i}(\mathbf{b}|\eta) = \log \widetilde{L}_{m_i}(\mathbf{b}|\eta)$. We denote a multivariate normal density by $\varphi(\cdot)$ and a unit vector in R^q by $\mathbf{u}$.

Lemma A.1

Assume conditions C.2–C.5. For all $1 \leq i \leq n$, given $\mathbf{b}_i$,

- i. $\|m_i^{-1/2} \sum_{j=1}^{m_i} \mathbf{z}(\mathbf{x}_{ij}, y_{ij}, \widehat{\mathbf{b}}_i)\| = o_p(1)$.
- ii. $\left\| \sum_{j=1}^{m_i} \{\mathbf{z}(\mathbf{x}_{ij}, y_{ij}, \mathbf{b}) - E\mathbf{z}(\mathbf{x}_{ij}, y_{ij}, \mathbf{b})\} \right\| = O_p(m_i^{1/2})$ and $m_i^{-1} \sum_{j=1}^{m_i} \mathbf{z}(\mathbf{x}_{ij}, y_{ij}, \mathbf{b}) \mathbf{z}^\top(\mathbf{x}_{ij}, y_{ij}, \mathbf{b}) = \mathbf{D}_{i0} + o_p(1)$, uniformly for all $\mathbf{b}$ in an $o(1)$ -neighborhood of $\mathbf{b}_i$.
- iii. $m_i^{-1/2} \sum_{j=1}^{m_i} \mathbf{z}(\mathbf{x}_{ij}, y_{ij}, \mathbf{b}) = m_i^{-1/2} \sum_{j=1}^{m_i} \mathbf{z}(\mathbf{x}_{ij}, y_{ij}, \mathbf{b}_i) - \mathbf{D}_{i1} \{m_i^{1/2}(\mathbf{b} - \mathbf{b}_i)\} + O_p(m_i^{-1/2})$, uniformly for all $\mathbf{b}$ in an $O(m_i^{-1/2})$ -neighborhood of $\mathbf{b}_i$.

Proof

(i) follows from Lemma A.2 in Ruppert and Carroll (1980) and (ii)–(iii) follows from Lemma A.5 in Yang and He (2011).

Lemma A.2

Assume conditions C.2–C.5. Given $\mathbf{b}_i$, there exists $\varepsilon_1 > 0$ for any δ_1 such that

$$\liminf_{m_i \rightarrow \infty} P \left\{ \sup_{\|\mathbf{b} - \mathbf{b}_i\| \geq \delta_1} m_i^{-1} \{l_{m_i}(\mathbf{b}) - l_{m_i}(\mathbf{b}_i)\} \leq -\varepsilon_1 \right\} = 1. \quad (\text{A.1})$$

Also for $\mathbf{b}$ in an open neighborhood of $\mathbf{b}_i$, $l_{m_i}(\mathbf{b}) = -\frac{m_i}{2}(\mathbf{b} - \widehat{\mathbf{b}}_i)^\top \mathbf{Q}_i^{-1}(\mathbf{b} - \widehat{\mathbf{b}}_i) + R_{m_i}(\mathbf{b})$, where there exists a sufficiently small $\delta_2 > 0$ and large $M > 0$ for each $\varepsilon_2 > 0$ such that

$$\begin{aligned} \limsup_{m_i \rightarrow \infty} P \left\{ \sup_{m_i^{-1/2} M \leq \|\mathbf{b} - \mathbf{b}_i\| \leq \delta_2} (m_i \|\mathbf{b} - \mathbf{b}_i\|^2)^{-1} |R_{m_i}(\mathbf{b})| > \varepsilon_2 \right\} &< \varepsilon_2, \\ \limsup_{m_i \rightarrow \infty} P \left\{ \sup_{\|\mathbf{b} - \mathbf{b}_i\| \leq m_i^{-1/2} M} |R_{m_i}(\mathbf{b})| > \varepsilon_2 \right\} &= 0. \end{aligned} \quad (\text{A.2})$$

Proof

With $\mathbf{b}_i$ replacing $\widehat{\mathbf{b}}_i$, the above results follow directly from Lemma A.2 and A.4 in Yang and He (2011) or Proposition 2 in Chernozhukov and Hong (2003). The results still holds with $\widehat{\mathbf{b}}_i$ by Lemma A.1-(i).

Lemma A.3

Assume conditions C.1 and C.1.(i) for $g(\mathbf{b}|\boldsymbol{\eta})$. We define $\{\boldsymbol{\Lambda}(\boldsymbol{\eta})\}^{-1} = -\{\partial^2 \log g / (\partial \mathbf{b} (\partial \mathbf{b})^\top)\}_{\mathbf{b}=\beta}$ and $\tilde{\mathbf{Q}}_i(\boldsymbol{\eta}) = [\mathbf{Q}_i^{-1} + m_i^{-1} \{\boldsymbol{\Lambda}(\boldsymbol{\eta})\}^{-1}]^{-1}$. Also assume conditions C.2–C.5. If $(\min_{1 \leq i \leq n} m_i) n^{-1/2} \rightarrow \infty$, for any $\boldsymbol{\eta}$ in an $o(1)$ -neighborhood of $\boldsymbol{\eta}_0$,

$$\log \tilde{\omega}_n(\boldsymbol{\eta}) = \sum_{i=1}^n \log g(\widehat{\mathbf{b}}_i|\boldsymbol{\eta}) + R_n(\boldsymbol{\eta}), \quad (\text{A.3})$$

where similar results as in (A.2) hold for $R_n(\boldsymbol{\eta})$.

Proof

Under C.1.(i) it holds that for any $\mathbf{b}, \mathbf{b}^* \in \Upsilon$ with $\|\mathbf{b} - \mathbf{b}^*\| < \varepsilon$,

$$\log g(\mathbf{b}|\boldsymbol{\eta}) = \log g(\mathbf{b}^*|\boldsymbol{\eta}) - \frac{1}{2}(\mathbf{b} - \beta)^\top \{\boldsymbol{\Lambda}(\boldsymbol{\eta})\}^{-1}(\mathbf{b} - \beta) + \frac{1}{2}(\mathbf{b}^* - \beta)^\top \{\boldsymbol{\Lambda}(\boldsymbol{\eta})\}^{-1}(\mathbf{b}^* - \beta) + (\mathbf{b} - \mathbf{b}^*)^\top \mathbf{r}(\mathbf{b}, \mathbf{b}^*, \boldsymbol{\eta}), \quad (\text{A.4})$$

where $\|E\{\mathbf{r}(\mathbf{b}, \mathbf{b}^*, \boldsymbol{\eta})\}\| < \infty$ and $E[\mathbf{u}^\top \{\partial \mathbf{r}(\mathbf{b}, \mathbf{b}^*; \boldsymbol{\eta}) / \partial \boldsymbol{\eta}\} \mathbf{u}] < \infty$ for all $\boldsymbol{\eta} \in \Psi$ as $\varepsilon \rightarrow 0$. As $\|\mathbf{b}_i - \widehat{\mathbf{b}}_i\| = O_p(m_i^{-1/2})$, it follows from Lemma A.2 and (A.4) that given $\mathbf{b}_i$ and for all $\mathbf{b}$ in a $o(1)$ -neighborhood of $\mathbf{b}_i$,

$$\widetilde{l}_{m_i}(\mathbf{b}|\boldsymbol{\eta}) = v_{m_i}(\mathbf{b}, \widehat{\mathbf{b}}_i, \boldsymbol{\eta}) + \rho_{m_i}(\widehat{\mathbf{b}}_i, \boldsymbol{\eta}), \quad (\text{A.5})$$

where $v_{m_i}(\mathbf{b}, \widehat{\mathbf{b}}_i, \boldsymbol{\eta}) = \log \varphi(\mathbf{b}|\widehat{\mathbf{b}}_i^*(\boldsymbol{\eta}), m_i^{-1} \tilde{\mathbf{Q}}_i(\boldsymbol{\eta})) + R_{m_i}(\mathbf{b}) + (\mathbf{b} - \mathbf{b}^*)^\top \mathbf{r}(\mathbf{b}, \mathbf{b}^*, \boldsymbol{\eta})$ and

$\rho_{m_i}(\widehat{\mathbf{b}}_i, \boldsymbol{\eta}) = \log g(\widehat{\mathbf{b}}_i|\boldsymbol{\eta}) + \frac{1}{2}(\widehat{\mathbf{b}}_i - \beta)^\top [\boldsymbol{\Lambda}(\boldsymbol{\eta})\{\tilde{\mathbf{Q}}_i(\boldsymbol{\eta})\}^{-1} \boldsymbol{\Lambda}(\boldsymbol{\eta})]^{-1}(\widehat{\mathbf{b}}_i - \beta) + \frac{1}{2} \log |m_i^{-1} \tilde{\mathbf{Q}}_i(\boldsymbol{\eta})|$, and $\widehat{\mathbf{b}}_i^*(\boldsymbol{\eta}) = \tilde{\mathbf{Q}}_i(\boldsymbol{\eta})[\mathbf{Q}_i^{-1} \widehat{\mathbf{b}}_i + m_i^{-1} \{\boldsymbol{\Lambda}(\boldsymbol{\eta})\}^{-1} \beta]$. As $(\min_{1 \leq i \leq n} m_i) n^{-1/2} \rightarrow \infty$ and (A.5) holds true for all $1 \leq i \leq n$ with $\mathbf{b} = \mathbf{b}_i$, we have

$$\log \tilde{\omega}_n(\boldsymbol{\eta}) = \sum_{i=1}^n \log g(\widehat{\mathbf{b}}_i|\boldsymbol{\eta}) + \sum_{i=1}^n \log U_i(\widehat{\mathbf{b}}_i, \boldsymbol{\eta}) + o_p(1),$$

where $U_i(\widehat{\mathbf{b}}_i, \boldsymbol{\eta}) = \int_{\mathbf{b}_i \in \Upsilon} \exp\{v_{m_i}(\mathbf{b}_i, \widehat{\mathbf{b}}_i, \boldsymbol{\eta})\} d\mathbf{b}_i$. Consider $\boldsymbol{\zeta}_i = m_i^{1/2} \{\tilde{\mathbf{Q}}_i(\boldsymbol{\eta})\}^{-1/2} \{\mathbf{b}_i - \widehat{\mathbf{b}}_i^*(\boldsymbol{\eta})\}$. For some appropriately modified functions $R_{m_i}^*(\boldsymbol{\zeta}_i, \widehat{\mathbf{b}}_i)$ and $\mathbf{r}^*(\boldsymbol{\zeta}_i, \widehat{\mathbf{b}}_i, \boldsymbol{\eta})$,

$$U_i(\widehat{\mathbf{b}}_i, \boldsymbol{\eta}) = \int \varphi(\boldsymbol{\zeta}_i|\mathbf{0}, \mathbf{I}) \exp[R_{m_i}^*(\boldsymbol{\zeta}_i, \widehat{\mathbf{b}}_i) + \{m_i^{-1/2} \tilde{\mathbf{Q}}_i^{1/2}(\boldsymbol{\eta}) \boldsymbol{\zeta}_i + (\widehat{\mathbf{b}}_i^*(\boldsymbol{\eta}) - \widehat{\mathbf{b}}_i)\}^\top \mathbf{r}^*(\boldsymbol{\zeta}_i, \widehat{\mathbf{b}}_i, \boldsymbol{\eta})] d\boldsymbol{\zeta}_i.$$

We note that $|U_i(\widehat{\mathbf{b}}_i, \boldsymbol{\eta})| = O_p(1)$ by Lemma A.2 for any $\boldsymbol{\eta} \in \Psi$ and $U_i(\widehat{\mathbf{b}}_i, \boldsymbol{\eta})$ is continuous in $\boldsymbol{\eta}$. Also as $\tilde{\mathbf{Q}}_i(\boldsymbol{\eta}) = \mathbf{b}_i - m_i^{-1} \mathbf{b}_i \{\boldsymbol{\Lambda}(\boldsymbol{\eta}) + m_i^{-1} \mathbf{b}_i\}^{-1} \mathbf{b}_i$, $\|\mathbf{u}^\top (\partial \boldsymbol{\zeta}_i / \partial \boldsymbol{\eta}) \mathbf{u}\| = O_p(m_i^{-1/2})$, and $\|\mathbf{u}^\top \{\partial (\widehat{\mathbf{b}}_i^*(\boldsymbol{\eta}) - \widehat{\mathbf{b}}_i) / \partial \boldsymbol{\eta}\} \mathbf{u}\| = O_p(m_i^{-1})$ for all $\boldsymbol{\eta} \in \Psi$, and $\int \{\partial \log \varphi(\boldsymbol{\zeta}_i|\mathbf{0}, \mathbf{I}) / \partial \boldsymbol{\eta}\} \varphi(\boldsymbol{\zeta}_i|\mathbf{0}, \mathbf{I}) d\boldsymbol{\zeta}_i = 0$.

It follows that $\left\| \partial \log U_i(\widehat{\mathbf{b}}_i, \eta) / \partial \eta \right\| = O_p(m_i^{-1})$ for all $\eta \in \Psi$ and all $1 \leq i \leq n$. The desired results follow as $n \rightarrow \infty$ with $(\min_{1 \leq i \leq n} m_i) n^{-1/2} \rightarrow \infty$.

Proof of Theorems

Proofs rely on the factorization of the likelihood like function $\widetilde{l}_{m_i}(\mathbf{b}_i|\eta)$ in an asymptotic sense. We sketch the proofs below.

Proof of Theorem 3

We prove the result (ii) first. We let

$$\Delta_{\eta_0} = \left\{ \partial \sum_{i=1}^n \log g(\widehat{\mathbf{b}}_i|\eta) / \partial \eta \right\}_{\eta=\eta_0}, \quad n\mathbf{J}_{\eta_0} = -E \left\{ \partial^2 \sum_{i=1}^n \log g(\widehat{\mathbf{b}}_i|\eta) / (\partial \eta \partial \eta^\top) \right\}_{\eta=\eta_0}.$$

By Lemma A.3 and expanding (A.3) around η_0 , for any η with $\|\eta - \eta_0\| = O(n^{-1/2})$,

$$\log \tilde{\omega}_n(\eta) - \log \tilde{\omega}_n(\eta_0) = (\eta - \eta_0)^\top \Delta_{\eta_0} - \frac{1}{2} (\eta - \eta_0)^\top (n\mathbf{J}_{\eta_0}) (\eta - \eta_0) + o_p(1). \quad (\text{A.6})$$

Under condition C.1.(i) we note that $\{Var(n^{-1/2}\Delta_{\eta_0})\}^{1/2} n^{-1/2}\Delta_{\eta_0} \rightarrow AN(\mathbf{0}, \mathbf{I})$ and $\mathbf{J}_{\eta_0}$ is positive definite almost surely as $n \rightarrow \infty$ since $(\min_{1 \leq i \leq n} m_i) n^{-1/2} \rightarrow \infty$. It follows from standard Bayesian equivalent of the central limit theorem results such as Bernardo and Smith (2003) (Chapter 5.3) that we have

$$n^{1/2}(\tilde{\eta}_n - \eta_0) \sim AN(\mathbf{0}, \mathbf{J}_{\eta_0}^{-1} \Omega_{\eta_0} \mathbf{J}_{\eta_0}^{-1}),$$

where $\Omega_{\eta_0} = Var(n^{-1/2}\Delta_{\eta_0}) + o_p(1)$. The result (ii) follows by the Delta method.

For the proof of (i) we note that similar result as in (A.1) in Lemma A.2 holds for $\log \tilde{\omega}_{ni}(\mathbf{b})$ under condition C.1.(i). As $|\mathbf{u}^\top (\partial \widehat{\mathbf{b}}_i^*(\eta) / \partial \eta) \mathbf{u}| = O(m_i^{-1})$ for all $\eta \in \Psi$, it follows from (A.5) and by the definition of $\omega_{ni}(\cdot)$ that for any $\mathbf{b}$ in an open neighborhood of $\mathbf{b}_i$ that is given,

$$\log \tilde{\omega}_{ni}(\mathbf{b}) = \log \varphi\{\mathbf{b}|\widehat{\mathbf{b}}_i^*(\eta_0), m_i^{-1}\tilde{\mathbf{Q}}_i(\eta_0)\} + \tilde{R}_{m_i}(\mathbf{b}), \quad (\text{A.7})$$

where results corresponding to (A.2) hold for $\tilde{R}_{m_i}(\mathbf{b})$. The rest of the proof follows by Theorem 2 of Chernozhukov and Hong (2003).

Proof of Theorem 1 and 2

We first prove the results assuming $\Sigma_i = \Sigma$. For the proof of Theorem 1, it suffices to note that in (A.4) $\mathbf{r}(\mathbf{b}, \mathbf{b}^*, \theta) = 0$ and $\{\Lambda(\eta)\}^{-1} = \Sigma^{-1}$ and (A.5) becomes

$$\begin{aligned}\widetilde{l}_{m_i}(\mathbf{b}|\theta) &= \log \varphi(\mathbf{b}|\widehat{\mathbf{b}}_i^*(\theta), m_i^{-1}\widetilde{\mathbf{Q}}_i(\theta)) + R_{m_i}(\mathbf{b}) \\ &\quad + \log \varphi(\widehat{\mathbf{b}}_i|\beta, \mathbf{V}_i(\theta)) + (1/2)\log|m_i^{-1}\mathbf{Q}_i|,\end{aligned}\tag{A.8}$$

where $\mathbf{V}_i(\theta) = \sum + m_i^{-1}\mathbf{Q}_i$. For (iii), we note that it follows from McCulloch (1982) that

$$E \left\{ - \frac{\partial^2 \varphi(\widehat{\mathbf{b}}_i|\beta, \mathbf{V}_i(\theta))}{\partial \text{vech}(\mathbf{V}_i(\theta))(\partial \text{vech}(\mathbf{V}_i(\theta)))^\top} \Big|_{\theta=\theta_0} \right\} = \frac{1}{2} \mathbf{K}^\top (\mathbf{V}_{i0}^{-1} \otimes \mathbf{V}_{i0}^{-1}) \mathbf{K},$$

where $\mathbf{K}$ defines the matrix operator $\text{vech}(\cdot)$ as shown in the theorem. The desired results follow as the Jacobians of $\text{vech}(\boldsymbol{\Sigma})$ with respect to $\text{vech}(\mathbf{V}_i)$ are identity matrices. On the other hand, the results (i) of Theorem 2 follows from Theorem 3 and 4 of Chernozhukov and Hong (2003). (ii) is given by standard Bayesian results respectively. For the proof when $\boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}$ for some i , it suffices to point that (A.8) remains true with $\boldsymbol{\Sigma}_i$.

Figure 1.

Trace plots of the Markov chains: the top panel contains the trace plots for $\beta(0.90)$, population averaged regression coefficients of a second degree polynomial at $\tau = 0.90$. The bottom panel contains the trace plots for corresponding random coefficients in the first cluster (or addiction treatment center).

Figure 2.

Scatter plots of days between first contact and first treatment against month of admission for four centers: m indicate per center sample sizes and super imposed curves represent regression fits by three different methods: REQR, REML and standard independent QR method.

Figure 3. Fitted second degree polynomials for the conditional quantiles at $\tau \in \{0.50, 0.75, 0.90\}$.

Figure 4.

Point and interval estimates by the proposed REQR and independent data QR method: the interval estimates correspond to 95% credible intervals for the REQR analysis and 95% confidence intervals for the per center analysis by the independent data QR method.

Figure 5.
The scatter of apnea duration times in seconds per subject on the log scale.

Figure 6. Boxplots of apnea duration times in seconds on the log scale: Wide boxes represent subjects being fed by the examiners, and narrow boxes represent self-feeding. Shaded boxes are pudding and white boxes are liquid. The horizontal lines are the sample quantiles for $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$.

Figure 7.

Trace plots of the Markov chains: the top panel contains the trace plots of the random intercepts and the bottom panel slopes respectively of 4 clusters in a simulation setting with homogeneous errors from t distribution with 2 degree of freedom. For each plot, the acceptance rates of Metropolis-Hastings algorithm are reported in the title.

Table 1

95% interval estimates for the apnea duration example: the REQR method is compared with the standard QR method that assumes independency for the within subject correlation. Intervals with asterisks do not included 0.

	τ	REQR	QR	Relative Length (REQR/QR)
Intercept	0.5	(-0.21, -0.04) *	(-0.18, -0.09) *	1.89
Viscosity (Visc)	0.5	(0.05, 0.38) *	(0.07, 0.26) *	1.74
Feed Cond (FC)	0.5	(-0.07, 0.04)	(-0.09, 0.05)	0.79
Volume (Vol)	0.5	(-0.03, 0.04)	(-0.06, 0.09)	0.47
Visc $\times$ FC	0.5	(0.04, 0.20) *	(0.02, 0.20) *	0.89
Visc $\times$ Vol	0.5	(-0.09, -0.02) *	(-0.22, -0.03) *	0.37
FC $\times$ Vol	0.5	(-0.02, 0.05)	(-0.06, 0.12)	0.39
Intercept	0.9	(0.02, 0.36) *	(0.35, 0.70) *	0.97
Viscosity (Visc)	0.9	(0.10, 0.50) *	(0.20, 0.62) *	0.95
Feed Cond (FC)	0.9	(-0.14, 0.07)	(-0.35, 0.17)	0.40
Volume (Vol)	0.9	(-0.04, 0.10)	(-0.32, 0.18)	0.28
Visc $\times$ FC	0.9	(0.16, 0.56) *	(0.34, 0.86) *	0.77
Visc $\times$ Vol	0.9	(-0.28, -0.03) *	(-0.38, 0.18)	0.45
FC $\times$ Vol	0.9	(-0.02, 0.12)	(-0.24, 0.35)	0.24

Table 2

Monte-Carlo study results for $\tau = 0.5$ with $n = 20$ and $m = 10$ based on 300 replicated datasets: MSE^{**} indicates the ratio of mean squared errors (MSE) for REML method over MSE for the REQR method. AL^{**} similarly indicates the ratio of average interval lengths. ASD is the average of standard error estimates and CP (%) is the coverage of 95% intervals in percent. N^{***} reports # of replicated datasets REML reported non-zero variance estimates. With REML, ASD and AL^{**} are computed only over the N^{***} # of cases with non-zero variance estimates. Numbers inside parentheses indicate standard errors of respective quantities.

e_{ij}		REML/REQR				N^{***}
		$MSE^{**}(REML/REQR)$	ASD	$CP(\%)$	$AL^{**}(REML/REQR)$	
LN	β_0	0.56 (0.222/0.284)	0.389 (0.083)/0.318 (0.086)	93.3/71.3	1.23 (0.33/0.34)	
	β_1	0.94 (0.105/0.136)	0.276 (0.046)/0.266 (0.050)	92.0/90.0	1.04 (0.18/0.20)	
	$\sum^2$	1.58 (1.609/2.308)	0.870 (0.361)/0.630 (0.359)	89.3/86.3	1.46 (1.42/1.38)	273
	$\sum_0^2$					
N	β_0	1.17 (0.362/0.629)	0.428 (0.143)/0.469 (0.202)	88.7/96.0	0.94 (0.56/0.76)	300
	β_1					
	$\sum^2$					
	$\sum_0^2$					
t_2	β_0	0.80 (0.096/0.110)	0.263 (0.045)/0.287 (0.069)	95.0/92.3	0.91 (0.18/0.28)	
	β_1	0.90 (0.071/0.090)	0.230 (0.037)/0.259 (0.041)	92.3/95.3	0.89 (0.14/0.16)	
	$\sum^2$	0.78 (0.399/0.545)	0.450 (0.160)/0.547 (0.274)	87.7/86.7	0.86 (0.63/1.05)	300
	$\sum_0^2$	0.53 (0.180/0.236)	0.343 (0.114)/0.470 (0.144)	87.3/92.3	0.75 (0.45/0.55)	300
$z_{ij} \times LN$	β_0	1.31 (0.644/0.222)	0.474 (0.295)/0.345 (0.102)	95.0/90.7	1.37 (1.16/0.40)	
	β_1	1.45 (0.527/0.111)	0.314 (0.171)/0.283 (0.048)	92.0/92.0	1.11 (0.67/0.19)	
	$\sum^2$	2.32 (4.763/1.402)	1.103 (0.635)/0.751 (0.491)	75.7/87.0	1.56 (2.49/1.60)	231
	$\sum_0^2$	31.96 (148.7/0.335)	0.569 (1.602)/0.522 (0.183)	89.3/92.7	1.12 (6.28/0.70)	296
$z_{ij} \times LN$	β_0	1.51 (0.336/0.225)	0.578 (0.185)/0.340 (0.092)	97.7/90.0	1.71 (0.72/0.36)	
	β_1	0.66 (0.254/0.358)	0.402 (0.098)/0.318 (0.070)	93.3/78.7	1.26 (0.39/0.27)	

e_{ij}	MSE^*	$(REML/REQR)$	REML/REQR		CP(%)	AL^{**}	$(REML/REQR)$	N^{***}
			ASD					
$\sum_0^2$	1.81	(0.212/0.910)	1.204	(0.542/0.701	11.7/86.0	1.83	(2.16/1.41)	35
$\sum_1^2$	1.77	(1.678/0.799)	0.658	(0.309/0.637	96.3/87.3	1.08	(1.21/1.00)	299
β_0	0.70	(0.150/0.149)	0.313	(0.038/0.303	95.3/85.0	1.03	(0.15/0.31)	
β_1	0.78	(0.096/0.128)	0.283	(0.040/0.296	97.0/94.0	0.95	(0.16/0.22)	
$\sum_0^2$	1.28	(0.376/0.759)	0.600	(0.162/0.585	56.3/78.0	1.08	(0.64/1.15)	193
$\sum_1^2$	0.71	(0.448/0.534)	0.449	(0.148/0.561	96.0/94.0	0.84	(0.58/0.80)	300
β_0	1.95	(0.777/0.212)	0.739	(0.365/0.350	98.7/89.0	2.12	(1.43/0.46)	
β_1	1.52	(0.635/0.209)	0.483	(0.194/0.335	95.0/92.0	1.44	(0.76/0.29)	
$\sum_0^2$	2.27	(4.419/0.713)	3.042	(4.661/0.723	6.0/85.0	4.48	(18.3/1.62)	18
$\sum_1^2$	11.72	(67.97/0.802)	0.909	(1.143/0.690	94.7/90.7	1.38	(4.48/1.11)	296

Table 3

Monte-Carlo study results for $\tau = 0.5$ with $n = 20$ and $m = 20$ based on 500 simulations: MSE^* and AL^{**} report the ratios of mean squared errors (MSE) and average interval lengths as in Table 2. CP (%) is the coverage of 95% intervals in percent.

	e_{ij}	MSE^*	CP (%)			e_{ij}	MSE^*	CP (%)			AL^{**}	AL^{**}
			RERQ	REML	AL^{**}			RERQ	REML	AL^{**}		
β_0	β_0	0.769	86.5	94.2	1.145	e_{ij}	1.509	92.6	97.0	1.485	AL^{**}	1.485
		0.972	95.6	94.8	0.982		0.711	81.8	94.4	1.169		1.169
β_1	β_1	1.830	90.4	89.2	1.250	$z_{ij} \times LN$	2.919	91.8	25.8	1.599	AL^{**}	1.599
$\sum_{i=0}^2$	$\sum_{i=0}^2$	0.952	94.8	87.7	0.876	$z_{ij} \times N$	1.421	94.4	97.2	1.012	AL^{**}	1.012
β_0	β_0	0.928	91.8	92.2	0.943	$z_{ij} \times N$	0.909	93.2	93.6	0.948	AL^{**}	0.948
		0.987	95.4	93.6	0.925		0.885	96.4	96.2	0.930		0.930
β_1	β_1	0.629	88.6	90.0	0.856	$z_{ij} \times N$	1.432	92.3	69.1	0.872	AL^{**}	0.872
$\sum_{i=0}^2$	$\sum_{i=0}^2$	0.748	94.4	88.0	0.813	$z_{ij} \times N$	0.734	94.0	94.5	0.806	AL^{**}	0.806
β_0	β_0	3.195	93.2	93.0	1.356	$z_{ij} \times t_2$	3.532	93.2	98.4	1.963	AL^{**}	1.963
		1.639	94.8	92.6	1.075		3.070	94.0	96.0	1.360		1.360
β_1	β_1	2.718	87.9	83.6	1.620	$z_{ij} \times t_2$	3.254	89.0	17.4	2.669	AL^{**}	2.669
$\sum_{i=0}^2$	$\sum_{i=0}^2$	1.871	92.8	89.2	0.954	$z_{ij} \times t_2$	130.226	91.4	94.6	1.509	AL^{**}	1.509

Table 4
Coverage (%) of 95% intervals for $\tau = 0.9$ based on 300 simulations with $n = 20$.

e_{ij}	LN			N			t_2		
	m	20	50	20	50	20	20	50	20
β_0		75.7	89.3	86.3	93.7	83.7	83.7	91.3	
β_1		89.0	82.0	94.3	94.7	89.3	89.3	88.7	
σ_0^2		82.7	86.3	94.0	93.3	84.7	84.7	89.3	
σ_1^2		94.0	94.3	94.3	95.7	94.7	94.7	94.7	

e_{ij}	$z_{ij} \times LN$			$z_{ij} \times N$			$z_{ij} \times t_2$		
	m	20	100	20	100	20	20	100	20
β_0		44.0	42.3	66.7	80.3	50.3	50.3	54.3	
β_1		60.3	91.3	62.7	94.7	76.7	76.7	96.0	
σ_0^2		90.3	89.7	94.3	95.0	92.3	92.3	90.7	
σ_1^2		86.7	87.3	95.3	95.7	89.3	89.3	90.0	