1   Background

Mixed effects model methodology, first introduced by R.A. Fisher [1], is a statistical tool that is used across a wide variety of disciplines including biostatistical contexts. Mixed effects models are especially popular in researches involving repeated measurements or observations from multilevel (or hierarchical) structure where the correlation between observations is not negligible.

Quantile regression models are becoming more and more popular in the statistical community in recent years. Compared with the ubiquitous mean regression (a.n.a. linear regression) models, quantile regression models provide a much more comprehensive and focused insight into the association between the variables by studying the conditional quantiles functions of the outcome, which may not be observed by looking only at conditional mean models [5]. In quantile regression, the regression coefficients (β) are functions of the quantile (τ), and their estimates vary according to different quantiles. Thus quartile regression provides a way to studying the heterogeneity of the outcome that is associated with the covariates [5]. Moreover, quantile regression is more robust against outliers in the outcome compared with the mean regression, which is an immediate extension from the property of quantiles.

For a linear quantile regression model, [6] introduced a method in estimating the conditional quantiles. As an introductory material, Koenker [7] briefly covers the fundamentals of quantile regression, parameter estimation techniques, inference, asymptotic theory, etc., his book [5] provides more comprehensive and deeper introduction on quantile regression related topics. Yu and Moyeed [11] introduced the idea of Bayesian quantile regression by modeling the error term using asymmetric Laplace distribution (ALD) followed the idea proposed in [6]. [8] developed a Gibbs sampling algorithm for Bayesian quantile regression models, in which they used a location-scale mixture representation of the ALD. Many works have been done to extend the quantile regression method to accommodate longitudinal data. [4] developed a quansi-likehood method for median regression model for longitudinal data. [3] proposed to fit the quantile regression for longitudinal data based on ALD and the estimation is made by using Monte Carlo EM algorithm, later on [9] followed the idea of [3] and extended the model from random intercept to include random regression coefficients as well. [2] proposed a working correlation model for quantile regression for longitudinal data, a induced smoothing method was used to make the inference of the estimators. Fully Bayesian techniques were also developed for fitting linear quantile mixed models, for example [10] used the similar idea as [8] did in decomposing the error term as a location-scale mixture and developed a Gibbs sampling algorithm for quantile linear mixed model. The fully Bayesian method is appealing because it is easy to implement and to make inference, the uncertainty of the unknowns is taken into account, and it is flexible in the distribution of random effects. The detailed background about Bayesian quantile linear mixed model will be introduced in Section **??**. In addition, temporal

A novel idea that combines the linear quantile mixed model and nonparametric Bayesian methodology in studying the longitudinal data can be an important supplement to existing methods. Longitudinal or correlated data commonly arise in biostatistical context. For example, they can be measurements that are taken repeatedly over time, or outcomes from grouped patients in clinical trials. Moreover, real data can be far from ideal situation, which means extreme observations can be possibly included. It is crucial that a proper method to be used in order to make a better investigation of the relationship between the outcome of interest and potential risk factors. Traditional mean regression model is easy to grasp and implement, but it also suffers from its fragility in facing of the outliers. Ignoring the outliers in analysis could lead to misleading conclusion [7]. The quantile regression method can be a sensible supplement where the mean regression shouldn't be used in those situations.

[Add background for topic two here.]

To sum up, this research project focuses on developing and applying new statistical methods in analysis longitudinal data and will cover the following topics. First, a statistical method that combines nonparametric Bayesian technique and the quartile regression model in analyzing longitudinal data will be introduced. In this part, the linear quantile mixed model is used so that the within subject correlation is accounted and the

results also will be more robust to potential outlying observations. Since we used quantile regress instead of mean regression, a more comprehensive insight about the association between the outcome and covariates will be learned. As mentioned, in this part we will use DP to model the random effects in linear quantile mixed model, which is robust to the true underlying distribution of random effects and supposed to provide better prediction of the random effects. Second, as studying the temporal data, we will focus on the prediction of future outcomes.

References

[1] Ronald A Fisher. The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the royal society of Edinburgh*, 52(02):399–433, 1919.

[2] Liya Fu and You-Gan Wang. Quantile regression for longitudinal data with a working correlation model. *Computational Statistics & Data Analysis*, 56(8):2526–2538, 2012.

[3] Marco Geraci and Matteo Bottai. Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics*, 8(1):140–154, 2007.

[4] Sin-Ho Jung. Quasi-likelihood for median regression models. *Journal of the American Statistical Association*, 91(433):251–257, 1996.

[5] Roger Koenker. *Quantile regression*. Cambridge university press, 2005.

[6] Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

[7] Roger Koenker and Kevin Hallock. Quantile regression: An introduction. *Journal of Economic Perspectives*, 15(4):43–56, 2001.

[8] Hideo Kozumi and Genya Kobayashi. Gibbs sampling methods for bayesian quantile regression. *Journal of statistical computation and simulation*, 81(11):1565–1578, 2011.

[9] Yuan Liu and Matteo Bottai. Mixed-effects models for conditional quantiles with longitudinal data. *The International Journal of Biostatistics*, 5(1), 2009.

[10] Youxi Luo, Heng Lian, and Maozai Tian. Bayesian quantile regression for longitudinal data models. *Journal of Statistical Computation and Simulation*, 82(11):1635–1649, 2012.

[11] Keming Yu and Rana A Moyeed. Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447, 2001.