stats

MDPI

Article

Quantile Regression Approach for Analyzing Similarity of Gene Expressions under Multiple Biological Conditions

Dianliang Deng [1,*,†] and Mashfiqul Huq Chowdhury [2,†]

[1] Department of Mathematics and Statistics, University of Regina, Regina, SK S4S 0A2, Canada
[2] Department of Statistics, Mawlana Bhashani Science and Technology University, Santosh, Tangail 1902, Bangladesh; mashfiq@mbstu.ac.bd or chowdhma@uregina.ca
* Correspondence: deng@uregina.ca
† These authors contributed equally to this work.

Abstract: Temporal gene expression data contain ample information to characterize gene function and are now widely used in bio-medical research. A dense temporal gene expression usually shows various patterns in expression levels under different biological conditions. The existing literature investigates the gene trajectory using the mean function. However, temporal gene expression curves usually show a strong degree of heterogeneity under multiple conditions. As a result, rates of change for gene expressions may be different in non-central locations and a mean function model may not capture the non-central location of the gene expression distribution. Further, the mean regression model depends on the normality assumptions of the error terms of the model, which may be impractical when analyzing gene expression data. In this research, a linear quantile mixed model is used to find the trajectory of gene expression data. This method enables the changes in gene expression over time to be studied by estimating a family of quantile functions. A statistical test is proposed to test the similarity between two different gene expressions based on estimated parameters using a quantile model. Then, the performance of the proposed test statistic is examined using extensive simulation studies. Simulation studies demonstrate the good statistical performance of this proposed test statistic and show that this method is robust against normal error assumptions. As an illustration, the proposed method is applied to analyze a dataset of 18 genes in *P. aeruginosa*, expressed in 24 biological conditions. Furthermore, a minimum Mahalanobis distance is used to find the clustering tree for gene expressions.

Keywords: chi-square test; classification; linear mixed model; Mahalanobis distance; quantile analysis; temporal gene expressions

check for updates

Citation: Deng, D.; Chowdhury, M.H. Quantile Regression Approach for Analyzing Similarity of Gene Expressions under Multiple Biological Conditions. *Stats* **2022**, *5*, 583–605. https://doi.org/10.3390/stats5030036

Academic Editor: Zhu Wei

Received: 6 June 2022
Accepted: 29 June 2022
Published: 2 July 2022

1. Introduction

Recently, many researchers have focus on the analysis of gene expression data. The gene expression process records measurements of expression under various biological conditions over a specific time period. At present, micro-array experiments are widely being used to generate rapidly vast amounts of data on gene expression under various biological conditions. The analysis of temporal gene expression is now becoming of great interest to scientists to understand the complex mechanism of gene profiles and characterize gene expressions. Moreover, this analysis also helps bio-medical scientists detect the genes responsible for early cancer (Fang et al. [1]). Genes are generally expressed by transcription into RNA, and then this transcript might be translated into protein. Usually, the gene expression process conveys RNA information as numerical outcomes. For gene expression data, specific biological characteristics (for example, RNA information) are usually measured during a predetermined time interval under different experimental conditions.

Several statistical methods to analyze gene expression data have been considered, including clustering, fold expression changes, ANOVA, etc. Draghici and Kulaeva [2]

discussed a noise sampling method based on ANOVA for the selection of differentially regulated genes. They compared their results with the fold change method and discussed the risk of obtaining a false positive when strictly observing fold change. Eisen et al. [3] used a cluster analysis method to compare genes according to similarity. Li et al. [4] introduced a time-lagged correlation coefficient to assess the relationship between genes and used the linear mixed-effect model with splines for gene clustering. Yeung and Ruzzo [5] applied the principal component analysis to analyze gene expression data. They also studied the effectiveness of different clustering algorithms to capture cluster structure. Fang et al. [1] discussed some limitations of the fold expression method, suggesting that genetic information might be lost using this method. One defining feature of gene expression analysis is its time-dependency on the expression levels for a given gene at multiple times. Therefore, it is very important to incorporate the correlation structure within gene expressions. Kerr et al. [6], Storey and Tibshirani [7] studied differentially expressed genes from a single timepoint. Meanwhile, Tusher et al. [8] developed a temporal gene expression model based on one condition. Fang et al. [1] proposed a non-linear regression model to mark the relative change rates of genes. Their study uses changeable variance and covariance structure to model individual gene expression trajectory. Deng et al. [9] investigated the effects of different biological conditions on gene expressions. Under a given condition, Deng et al. [9] used log-normal distribution properties to characterize the variance function of genes and proposed a statistical test approach to test the equality of the variance function for different conditions. Deng et al. [10] studied the threshold points of the gene expressions and constructed test statistics to detect the threshold points.

At present, the available literature only considers using the mean function when analyzing gene expression trajectory. However, when gene expression data are skewed or over-dispersed, the mean function can be affected by outlying observations. Normality is usually assumed as the error term for estimation of the parameters in the model, and this assumption may be improper in practical instances. In addition, the temporal gene expression curve generally shows a strong degree of heterogeneity between multiple biological conditions. As a result, the rates of change for gene expressions may be different in non-central locations, and the mean model can only characterize the central location of the gene expression distribution. Thus, mean function may not properly capture the gene expression trajectory. However, by fitting the gene expression curves at different quantile vaules, the quantile regression approach can be used to completely examine the gene expression rate. Many researchers have explored statistical methods for the analysis of the quantile regression model. Huang and Lee [11] predicted the quantiles of daily Standard&Poor's 500 (S&P 500) returns by incorporating the high-frequency information through combining forecasts into one model. Most recently, Gallardo et al. [12] proposed a parametric quantile regression model for asymmetric response variables. Jung et al. [13] applied the multiple quantile regression method to the estimation of the spatial distribution of soil moisture. Chen et al. [14] studied estimation and inference for linear quantile regression models with generated regressors using a practical, two-step estimation procedure. Nevertheless, to our knowledge, there is no literature examining the gene expression data using the quantile regression model.

The purpose of this research is to apply the quantile regression (QR) model to the analysis of gene expression data and propose a test statistic for the examination of the similarity between gene expressions by comparing the quantile regression coefficients. The remainder of this article is organized as follows. In Section 2, the quantile regression model is proposed for gene expression data and the statistical inference is given for this model. The simulation study is performed in Section 3. An application for gene expression data is presented in Section 4, with the discussion in Section 5.

2. Quantile Model for Gene Expression Data

In this section, this research methodology is introduced to analyze the gene expression data. In most cases, for gene expression data, specific characteristics were measured at

discrete timepoints and measurements were taken under different biological conditions. As gene expression characteristics were measured over a series of timepoints and repeated under different circumstances, the longitudinal data analysis method can be applied to analyze discrete gene expression data. On the other hand, gene expression measurements show a variety of patterns under different biological conditions over specific times. Therefore, the linear quantile mixed model may be appropriate for the analysis of gene expression data. Let $Y_i(t)$ be the observed measurement of specific gene expression under the ith condition at time t. To determine the gene trajectory at different quantiles τ $(0 < \tau < 1)$, the following quantile model for $Y_i(t)$ can be considered.

$$\boldsymbol{Y}_i(t) = Q^{(\tau)}(t) + \boldsymbol{\epsilon}_i^{(\tau)}(t); \;\; i = 1, \ldots, N, \tag{1}$$

where, $Q^{(\tau)}(t)$ is the quantile curve at time t, $\epsilon_i^{(\tau)}(t)$ is the random noise with zero τth quantile and N denotes the number of conditions. Note that we are unable to observe $Y_i(t)$ for all timepoints, only at the very specific occasions t_{ij} at which measurements with errors were taken. Therefore, the observed longitudinal data for a specific gene under ith biological condition consist of the measurements $y_i = \{y_i(t_{ij}); \;\; j = 1, \cdots k_i\}$, where k_i denotes the number of observed timepoints. Note that, for gene expression data, there is no available covariate. However, there are many methods in the literature that can be used to estimate the quantile function $Q^{(\tau)}(t)$, including kernel, local polynomial, smoothing splines, regression splines and wavelet-based methods, among others. One simple and straightforward basis is the polynomial basis $\{1, t, \ldots, t^{p-1}\}$, in which the true quantile function of gene trajectory expression is modeled as polynomials of degree $p-1$. In this article, the method proposed in Donoho and Johnstone [15] and further developed in Zhang [16] is adopted. In this spline method, $Q^{(\tau)}(t)$ is approximated using the linear combination of a set of truncated power basis functions. Given a sequence of K interior knots $0 < \kappa_1 < \kappa_2 < \cdots < \kappa_K < T$ where T is the end time of observations, the regression spline basis functions of order p are $1, t, t^2, \ldots, t^p, (t-\kappa_1)_+^p, \ldots, (t-\kappa_K)_+^p$. Denoting the vector of $r(= K+1+p)$ basis functions by

$$B(t) = (1, t, t^2, \ldots, t^p, (t-\kappa_1)_+^p, \ldots, (t-\kappa_K)_+^p)^\top,$$

the regression spline smoothing is used to model $Q^{(\tau)}(t)$ using the linear combination of the basis functions $B(t)$, and the linear quantile mixed model can be written as

$$\boldsymbol{Y}_i(t) = [B(t)]^\top \boldsymbol{\beta}^{(\tau)} + [Z(t)]^\top \boldsymbol{U}^{(\tau)} + \boldsymbol{\epsilon}_i^{(\tau)}(t); \tag{2}$$

where the basis function for the fixed effects parameter is denoted as $B(t) = (1, t, t^2, \ldots, t^q, (t-\kappa_1)_+^q, \ldots, (t-\kappa_K)_+^q)^\top$ and $Z(t)$ is considered as basis function for random effects parameter, which could be the q-dimensional sub-vector of $B(t)(q \leq r)$. In particular, basis functions for the random intercept model and random slope model can be written as follows:

- Random intercept model, $Z(t) = (1)$;
- Random slope model $Z(t) = (1, t)^\top$.

For model formulation, gene expression data are considered longitudinal data in the form $\{Y_i(t_{ij}), B(t_{ij})^\top, Z(t_{ij})^\top\}$, for n biological conditions and k_i time occasions, i.e., $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, k_i$. Now, define

$$\boldsymbol{Y}_i = \begin{pmatrix} Y_i(t_{i1}) \\ Y_i(t_{i2}) \\ \vdots \\ Y_i(t_{ik_i}) \end{pmatrix}, \;\; \boldsymbol{B}_i = \begin{pmatrix} B^\top(t_{i1}) \\ B^\top(t_{i2}) \\ \vdots \\ B^\top(t_{ik_i}) \end{pmatrix}, \;\; \boldsymbol{Z}_i = \begin{pmatrix} Z^\top(t_{i1}) \\ Z^\top(t_{i2}) \\ \vdots \\ Z^\top(t_{ik_i}) \end{pmatrix}, \;\; \boldsymbol{\epsilon}_i^{(\tau)} = \begin{pmatrix} \epsilon_{i1}^{(\tau)} \\ \epsilon_{i2}^{(\tau)} \\ \vdots \\ \epsilon_{ik_i}^{(\tau)} \end{pmatrix}$$

and $\boldsymbol{\beta}^{(\tau)} = (\beta_1^{(\tau)}, \ldots, \beta_r)^\top$, $\boldsymbol{U}_i^{(\tau)} = (U_{i1}^{(\tau)}, \ldots, U_{iq}^{(\tau)})^\top$. Then, we have the matrix form of the model (2) as follows:

$$\boldsymbol{Y}_i = \boldsymbol{B}_i\boldsymbol{\beta}^{(\tau)} + \boldsymbol{Z}_i\boldsymbol{U}_i^{(\tau)} + \boldsymbol{\epsilon}_i^{(\tau)}$$

Now, for a longitudinal setup, assuming that $\boldsymbol{Y}_i$ conditioned on random effects, $\boldsymbol{U}_i$ are independently distributed and this conditional distribution follows the asymmetric Laplace (AL) distribution with location parameter $\boldsymbol{\mu}_i^{(\tau)} = \boldsymbol{B}_i\boldsymbol{\beta}^{(\tau)} + \boldsymbol{Z}_i\boldsymbol{U}_i^{(\tau)}$ and scale parameter $\boldsymbol{\sigma}^{(\tau)}$, respectively. Therefore, it can be written as

$$\boldsymbol{Y}_i|\boldsymbol{U}_i^{(\tau)} \sim AL(\boldsymbol{B}_i\boldsymbol{\beta}^{(\tau)} + \boldsymbol{Z}_i\boldsymbol{U}_i^{(\tau)}, \boldsymbol{\sigma}^{(\tau)}, \tau),$$

where $\boldsymbol{\beta}^{(\tau)}$ denotes the $r \times 1$ vector of fixed-effect parameters and τ is considered the quantile level. Assume that random effect vectors $\boldsymbol{U}_i^{(\tau)}$ are zero τ-quantile vectors and independent of the error term $\boldsymbol{\epsilon}_i^{(\tau)}$ of the model ($\boldsymbol{U}_i^{(\tau)} \perp \boldsymbol{\epsilon}_i^{(\tau)}$). Moreover, assume that random effect vectors $\boldsymbol{U}_i^{(\tau)}$ are distributed with the density function $f(\boldsymbol{u}_i^{(\tau)}|\boldsymbol{\Psi}^{(\tau)})$, where $\boldsymbol{\Psi}^{(\tau)}$ is regarded as the variance–covariance matrix (symmetric positive definite). All parameters depend on the skewness parameter, τ $(0 < \tau < 1)$. Now, the τth linear quantile mixed model can be written as

$$\boldsymbol{Y} = \boldsymbol{\mu}^{(\tau)} + \boldsymbol{\epsilon}^{(\tau)}, \tag{3}$$

where $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \boldsymbol{Y}_2^\top, \ldots, \boldsymbol{Y}_n^\top)^\top, \boldsymbol{\mu}^{(\tau)} = (\boldsymbol{\mu}_1^{(\tau)\top}, \boldsymbol{\mu}_2^{(\tau)\top}, \cdots, \boldsymbol{\mu}_n^{(\tau)\top})^\top$ and i.i.d components of error $\boldsymbol{\epsilon}^{(\tau)} = (\boldsymbol{\epsilon}_1^{(\tau)\top}, \boldsymbol{\epsilon}_2^{(\tau)\top}, \cdots, \boldsymbol{\epsilon}_n^{(\tau)\top})^\top$ follow an asymmetric Laplace distribution. Symbolically, $\epsilon_{ij}^{(\tau)} \sim AL(0, \sigma, \tau), i = 1, 2, \ldots, n; j = 1, 2, \ldots, k_i$. Now, in terms of matrix notation, τth linear quantile of response $(\boldsymbol{Y})$, denoted as $\boldsymbol{\mu}^{(\tau)}$, can also be expressed as

$$\boldsymbol{\mu}^{(\tau)} = \boldsymbol{B}\beta^{(\tau)} + \boldsymbol{Z}_\oplus\boldsymbol{U}^{(\tau)}, \tag{4}$$

where, $\boldsymbol{U}^{(\tau)} = (\boldsymbol{U}_1^\top, \boldsymbol{U}_2^\top, \cdots, \boldsymbol{U}_n^\top)^\top, \boldsymbol{Z}_\oplus = \bigoplus_{i=1}^{n} \boldsymbol{Z}_i$ and $\boldsymbol{B} = (\boldsymbol{B}_1^\top, \boldsymbol{B}_2^\top, \cdots, \boldsymbol{B}_n^\top)^\top$.

2.1. Estimation of Parameters

Based on the above discussion, the joint density of $(\boldsymbol{Y}, \boldsymbol{U}^{(\tau)})$ can be written in terms of the τth quantile as follows:

$$\begin{aligned} f(\boldsymbol{y}, \boldsymbol{u}^{(\tau)}|\boldsymbol{\beta}^{(\tau)}, \boldsymbol{\Psi}^{(\tau)}, \boldsymbol{\sigma}^{(\tau)}) &= f(\boldsymbol{y}|\boldsymbol{\beta}^{(\tau)}, \boldsymbol{u}^{(\tau)}, \boldsymbol{\sigma}^{(\tau)}) f(\boldsymbol{u}^{(\tau)}|\boldsymbol{\Psi}^{(\tau)}) \\ &= \prod_{i=1}^{n} f(\boldsymbol{y}_i|\boldsymbol{\beta}^{(\tau)}, \boldsymbol{u}_i^{(\tau)}, \boldsymbol{\sigma}^{(\tau)}) f(\boldsymbol{u}_i^{(\tau)}|\boldsymbol{\Psi}^{(\tau)}) \end{aligned} \tag{5}$$

For the random intercept model $(q = 1)$, the design matrix for random effects can be written as $\mathbf{Z}_i = (1, 1, \cdots, 1)'; i = 1, 2, \cdots, n$. Let, R^q denote $q-$dimensional Euclidean space. Now, the marginal likelihood can be derived from Equation (5) and written as

$$L_i(\boldsymbol{\beta}^{(\tau)}, \boldsymbol{\sigma}^{(\tau)}, \boldsymbol{\Psi}^{(\tau)}|\boldsymbol{y}) = \int_{R^q} f(\boldsymbol{y}_i|\boldsymbol{\beta}^{(\tau)}, \boldsymbol{u}_i^{(\tau)}, \boldsymbol{\sigma}^{(\tau)}) f(\boldsymbol{u}_i^{(\tau)}|\boldsymbol{\Psi}^{(\tau)}) d\boldsymbol{u}_i \tag{6}$$

Moreover, the marginal log-likelihood function can also be written as

$$l_i(\boldsymbol{\beta}^{(\tau)}, \boldsymbol{\sigma}^{(\tau)}, \boldsymbol{\Psi}^{(\tau)}|\boldsymbol{y}) = \log L_i(\boldsymbol{\beta}^{(\tau)}, \boldsymbol{\sigma}^{(\tau)}, \boldsymbol{\Psi}^{(\tau)}|\boldsymbol{y}). \tag{7}$$

Since the distribution of $Y(t_{ij})$ is assumed to follow an asymmetric Laplace distribution, τth quantile of $Y(t_{ij})$ can be estimated using the asymmetric Laplace distribution with location parameter $\mu_{ij}^{(\tau)} = \mu^{(\tau)}(t_{ij}) = [B(t_{ij})]^\top\boldsymbol{\beta}^{(\tau)} + [Z(t_{ij})]^\top U_i^{(\tau)}$, common scale $\sigma^{(\tau)}$ parameter and known skew parameter (τ). To estimate parameters from joint density

$f(\boldsymbol{y}, \boldsymbol{u}^{(\tau)})$, expressed in Equation (5), it is important to compute the following integral, which is also known as the marginal density of $\boldsymbol{Y}_i$.

$$f(\boldsymbol{y}_i|\boldsymbol{\beta}^{(\tau)}, \boldsymbol{\sigma}^{(\tau)}, \boldsymbol{\Psi}^{(\tau)}) = \boldsymbol{\sigma}_{k_i}^{(\tau)} \int_{R^q} \exp\left\{-\frac{1}{\boldsymbol{\sigma}}\rho_\tau(\boldsymbol{y}_i - \boldsymbol{\mu}_i^{(\tau)})\right\} f(\boldsymbol{u}_i|\boldsymbol{\Psi}^{(\tau)}) d\boldsymbol{u}_i, \tag{8}$$

where, $\boldsymbol{\sigma}_{k_i}^{(\tau)} = \left[\frac{\tau(1-\tau)}{\boldsymbol{\sigma}}\right]^{k_i}$ and

$$\rho_\tau(\boldsymbol{y}_i - \boldsymbol{\mu}_i^{(\tau)}) = \sum_{j=1}^{k_i} \rho_\tau(\boldsymbol{y}_{ij} - \mu_{ij}^{(\tau)}) = \sum_{j=1}^{k_i} \rho_\tau(y(t_{ij}) - [B(t_{ij})]^\top \boldsymbol{\beta}^{(\tau)} - [Z(t_{ij})]^\top U_i^{(\tau)})$$

Now, the log-likelihood function for n conditions can be expressed as

$$l(\boldsymbol{\beta}^{(\tau)}, \boldsymbol{\sigma}^{(\tau)}, \boldsymbol{\Psi}^{(\tau)}|\boldsymbol{y}) = \sum_{i=1}^{n} \left[\log(\boldsymbol{\sigma}_{k_i}^{(\tau)}) + \log \int_{R^q} \exp\left\{-\frac{1}{\boldsymbol{\sigma}}\rho_\tau(\boldsymbol{y}_i - \boldsymbol{\mu}_i^{(\tau)})\right\} f(\boldsymbol{u}_i|\boldsymbol{\Psi}^{(\tau)}) d\boldsymbol{u}_i\right] \tag{9}$$

This numerical likelihood can be solved by applying Gaussian quadrature (Gauss–Hermite quadrature or Gauss–Laguerre quadrature) proposed by Geraci and Bottai [17]. Now, assuming normal random effects ($\boldsymbol{U}_i \sim N(0, \boldsymbol{\Psi}^{(\tau)})$) to Equation (9), the Gauss–Hermite quadrature can be applied to approximate the likelihood function with nodes $\nu_{m_1,\cdots,m_q} = (\nu_{m_1}, \cdots, \nu_{m_q})'$ and weights $w_{m_l}, l = 1, 2, \cdots, q$, respectively. Integer M determines the number of points over the real line for each of the q one-dimensional integrals. The covariance matrix of the random effects is reparameterized by parameters $\boldsymbol{\alpha}^{(\tau)}$, i.e., $\boldsymbol{\Psi}(\boldsymbol{\alpha}^{(\tau)})$, and parameters, characterized by $\boldsymbol{\beta}^{(\tau)}$ and $\boldsymbol{\alpha}^{(\tau)}$, are denoted by $\boldsymbol{\theta}^{(\tau)} = (\boldsymbol{\beta}^{(\tau)}, \boldsymbol{\alpha}^{(\tau)})^T$. Finally, Equation (9) leads to the following approximate likelihood.

$$l_{app}(\boldsymbol{\theta}^{(\tau)}, \boldsymbol{\sigma}^{(\tau)}|\boldsymbol{y}) = \sum_{i=1}^{n} \log\left\{\sum_{m_1=1}^{M} \cdots \sum_{m_q=1}^{M} f\left(\boldsymbol{y}_i|\boldsymbol{\beta}^{(\tau)}, \boldsymbol{\sigma}^{(\tau)}, [\boldsymbol{\Psi}'(\boldsymbol{\alpha}^{(\tau)})]^{\frac{1}{2}} \nu_{m_1,\cdots,m_q}\right) \prod_{l=1}^{q} w_{m_l}\right\}. \tag{10}$$

Geraci and Bottai [17] develop the gradient search (gs) and the derivative free (df) optimization algorithm to maximize likelihood function in Equation (10). This optimization starts with a parameter value and then searches the positive semi-line for a new parameter value where likelihood is larger. This algorithm works until the likelihood change is sufficiently small or less than the pre-specified tolerance (δ) level. This algorithm begins estimating by initializing $\boldsymbol{\beta}^{(\tau)} = \boldsymbol{\beta}_0^{(\tau)}; \boldsymbol{\alpha}^{(\tau)} = \boldsymbol{\alpha}_0^{(\tau)}; \boldsymbol{\sigma}^{(\tau)} = \boldsymbol{\sigma}_0^{(\tau)}$. The derivative-free optimization algorithm is similar to the gradient search method. This method alternates a loop for $\boldsymbol{\theta}^{(\tau)}$ and then updates $\boldsymbol{\sigma}^{(\tau)}$.

Now, from the approximate likelihood in Equation (10), the fixed effects parameter $\boldsymbol{\beta}^{(\tau)}$, the random effects parameter $\boldsymbol{\alpha}^{(\tau)}$ and the error term parameter $\sigma^{(\tau)}$ can be estimated by using the algorithm given in Geraci and Bottai [17]. Further, the estimate of quantile function $Q^{(\tau)}(t)$ can be written as follows

$$\hat{Q}^{(\tau)}(t) = [B(t)]^\top \hat{\boldsymbol{\beta}}^{(\tau)} + [Z(t)]^\top \hat{\boldsymbol{U}}^{(\tau)}$$

where $\hat{\boldsymbol{\beta}}^{(\tau)}$ is the estimate of $\boldsymbol{\beta}^{(\tau)}$ and $\hat{\boldsymbol{U}}^{(\tau)}$ is the estimated best linear predictor of $\boldsymbol{U}^{(\tau)}$, which can be expressed as (Geraci and Botai [17])

$$\hat{\boldsymbol{U}}^{(\tau)} = \hat{\boldsymbol{\Psi}}^{(\tau)} \boldsymbol{Z}^\top \hat{\boldsymbol{\Sigma}}^{-1} \{\boldsymbol{Y} - \boldsymbol{B}^\top \hat{\boldsymbol{\beta}}^{(\tau)} - \hat{E}(\boldsymbol{\epsilon}^{(\tau)})\} \tag{11}$$

where $\hat{\boldsymbol{\Psi}}^{(\tau)} = \boldsymbol{\Psi}(\hat{\boldsymbol{\alpha}}^{(\tau)})$, $\boldsymbol{Z} = (\boldsymbol{Z}_1^\top, \ldots, \boldsymbol{Z}_n^\top)^\top$, $\hat{\boldsymbol{\Sigma}} = \widehat{\text{cov}}(Y) = \boldsymbol{Z}\hat{\boldsymbol{\Psi}}^{(\tau)}\boldsymbol{Z}^\top + \widehat{\text{cov}}(\boldsymbol{\epsilon}^{(\tau)}) = \boldsymbol{Z}\hat{\boldsymbol{\Psi}}^{(\tau)}\boldsymbol{Z}^\top + \text{diag}(\widehat{\text{var}}(\epsilon_{ij}^{(\tau)}))$ with $\widehat{\text{var}}(\epsilon_{ij}^{(\tau)}) = \frac{\hat{\sigma}^{(\tau)}(\tau^2 + (1-\tau)^2)}{\tau^2(1-\tau)^2}, j = 1, \ldots, k_i; i = 1, \ldots, n$

and $\hat{E}(\epsilon^{(\tau)}) = (\hat{E}(\epsilon_{11}^{(\tau)}), \ldots, \hat{E}(\epsilon_{1k_1}^{(\tau)}); \ldots; \hat{E}(\epsilon_{n1}^{(\tau)}), \ldots, \hat{E}(\epsilon_{nk_n}^{(\tau)}))$ with $\hat{E}(\epsilon_{ij}^{(\tau)}) = \frac{\hat{\sigma}^{(\tau)}(1-2\tau)}{\tau(1-\tau)}$ for $j = 1, \ldots, k_i; i = 1, \ldots, n$. Furthermore, the asymptotic covariance matrix for $\hat{\boldsymbol{U}}^{(\tau)}$ can be derived as

$$\mathrm{cov}(\hat{\boldsymbol{U}}^{(\tau)}) = \hat{\boldsymbol{\Psi}}^{(\tau)} \boldsymbol{Z}^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{Z} \hat{\boldsymbol{\Psi}}^{(\tau)} \boldsymbol{Z}^\top \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{Z} \hat{\boldsymbol{\Psi}}^{(\tau)} \tag{12}$$

Asymptotically, $\hat{Q}^{(\tau)}(t) \sim N(Q^{(\tau)}(t),\ \hat{V}_{Q^{(\tau)}}(t))$ with

$$\hat{V}_{Q^{(\tau)}}(t) = \widehat{\mathrm{var}}(\hat{Q}^{(\tau)}(t)) = (B^\top(t), Z^\top(t)) \widehat{\mathrm{cov}}(\hat{\boldsymbol{\beta}}^{(\tau)}, \hat{\boldsymbol{U}}^{(\tau)}) (B^\top(t), Z^\top(t))^\top$$

From the asymptotic normality of $\hat{Q}^{(\tau)}(t)$, the approximate $(1-\alpha)100\%$ confidence interval for the quantile function $Q^{(\tau)}(t)$ can be constructed as follows:

$$\hat{Q}^{(\tau)}(t) \pm z_{\frac{\alpha}{2}} \sqrt{\hat{V}_{Q^{(\tau)}}(t)} \tag{13}$$

where $z_{\frac{\alpha}{2}}$ is the upper 100 $(1-\frac{\alpha}{2})\%$ percentile of standard normal distribution. Although the asymptotic covariance matrix for $\hat{\boldsymbol{U}}^{(\tau)}$ has a closed form (12), there is no expression for the covariance matrix of estimator $\hat{\boldsymbol{\beta}}^{(\tau)}$, and thus we are unable to find the expression for the estimate of covariance $\widehat{\mathrm{cov}}(\hat{\boldsymbol{\beta}}^{(\tau)}, \hat{\boldsymbol{U}}^{(\tau)})$. However, bootstrap is a very powerful method and can be used in the estimation of covariance matrix for estimators $(\hat{\boldsymbol{\beta}}^{(\tau)}, \hat{\boldsymbol{U}}^{(\tau)})$. Here, we use a block bootstrap approach to find the estimate for $\widehat{\mathrm{cov}}(\hat{\boldsymbol{\beta}}^{(\tau)}, \hat{\boldsymbol{U}}^{(\tau)})$.

The procedures are as follows.

1. Obtain R bootstrap samples from the original data $\{Y_i(t_{ij}), B(t_{ij}), Z(t_{ij}); j = 1, 2, \ldots, k_i; i = 1, 2, \ldots, n\}$
2. Find the estimated values for the parameters $\boldsymbol{\beta}^{(\tau)}, \boldsymbol{\alpha}^{(\tau)}$ and $\boldsymbol{\sigma}^{(\tau)}$ and then calculate the values of $\hat{\boldsymbol{U}}^{(\tau)}$ by using formula (11) from each bootstrap sample and denote the obtained values as $\hat{\boldsymbol{\beta}}_1^{(\tau)}, \ldots, \hat{\boldsymbol{\beta}}_R^{(\tau)}; \hat{\boldsymbol{\alpha}}_1^{(\tau)}, \ldots, \hat{\boldsymbol{\alpha}}_R^{(\tau)}; \hat{\boldsymbol{\sigma}}_1^{(\tau)}, \ldots, \hat{\boldsymbol{\sigma}}_R^{(\tau)}$ and $\hat{\boldsymbol{U}}_1^{(\tau)}, \ldots, \hat{\boldsymbol{U}}_R^{(\tau)}$.
3. Set $\hat{\boldsymbol{\phi}}_r^{(\tau)} = (\hat{\boldsymbol{\beta}}_r^{(\tau)\top}, \hat{\boldsymbol{U}}_r^{(\tau)\top})^\top, \hat{\boldsymbol{\vartheta}}_r^{(\tau)} = (\hat{\boldsymbol{\beta}}_r^{(\tau)\top}, \hat{\boldsymbol{\alpha}}_r^{(\tau)}, \hat{\boldsymbol{\sigma}}_r^{(\tau)})^\top, r = 1, 2, \ldots, R$ and calculate the sample means of R bootstrap estimates for fixed effects parameters and random effects predictors $\boldsymbol{\phi}^{(\tau)} = (\boldsymbol{\beta}^{(\tau)\top}, \boldsymbol{U}^{(\tau)\top})^\top$ and $\boldsymbol{\vartheta}_r^{(\tau)} = (\boldsymbol{\beta}_r^{(\tau)\top}, \boldsymbol{\alpha}_r^{(\tau)}, \boldsymbol{\sigma}_r^{(\tau)})^\top$

$$\bar{\boldsymbol{\phi}}^{(\tau)} = \frac{1}{R} \sum_{r=1}^{R} \hat{\boldsymbol{\phi}}_r^{(\tau)}, \bar{\boldsymbol{\vartheta}}^{(\tau)} = \frac{1}{R} \sum_{r=1}^{R} \hat{\boldsymbol{\vartheta}}_r^{(\tau)} \tag{14}$$

4. Now, the bootstrap estimator for covariance matrix of estimators $(\hat{\boldsymbol{\beta}}^{(\tau)}, \hat{\boldsymbol{U}}^{(\tau)})$ can be written as

$$\hat{V}_{\boldsymbol{\phi}^{(\tau)}} = \widehat{\mathrm{cov}}(\hat{\boldsymbol{\beta}}^{(\tau)}, \hat{\boldsymbol{U}}^{(\tau)}) = \frac{1}{R-1} \sum_{r=1}^{R} (\hat{\boldsymbol{\phi}}_r^{(\tau)} - \bar{\boldsymbol{\phi}}^{(\tau)})(\hat{\boldsymbol{\phi}}_r^{(\tau)} - \bar{\boldsymbol{\phi}}^{(\tau)})^\top, \tag{15}$$

$$\hat{V}_{\boldsymbol{\vartheta}^{(\tau)}} = \widehat{\mathrm{cov}}(\boldsymbol{\beta}_r^{(\tau)\top}, \boldsymbol{\alpha}_r^{(\tau)}, \boldsymbol{\sigma}_r^{(\tau)}) = \frac{1}{R-1} \sum_{r=1}^{R} (\hat{\boldsymbol{\vartheta}}_r^{(\tau)} - \bar{\boldsymbol{\vartheta}}^{(\tau)})(\hat{\boldsymbol{\vartheta}}_r^{(\tau)} - \bar{\boldsymbol{\vartheta}}^{(\tau)})^\top \tag{16}$$

Furthermore, the estimates of model parameters and covariance matrix of estimators are computed by using the '**lqmm**' package developed by Geraci and Bottai [17] in the **R** programming environment. Generally, this '**lqmm**' package is used to estimate conditional quantile functions with random effects in linear quantile mixed models.

2.2. Test of the Similarity of Quantile Functions for Two Gene Expressions

The other purpose of this research is to identify gene similarity based on quantile functions. Using the results obtained in Section 2.1, the estimates of quantile functions

for gene expressions can be obtained. The estimated quantile functions for g genes can be expressed as

$$\hat{Q}_h^{(\tau)}(t) = [B(t)]^\top \hat{\boldsymbol{\beta}}_h^{(\tau)} + [Z(t)]^\top \hat{\boldsymbol{U}}_h^{(\tau)}; \;\; h = 1, 2, \cdots, g.$$

Two genes, h and s, are said to be similar if their quantile curve expressions are equal, i.e.,

$$\hat{Q}_h^{(\tau)}(t) = \hat{Q}_s^{(\tau)}(t); \;\; t \in [0, T], \;\; T > 0.$$

Therefore, the proposed hypothesis is

$$H_0 : Q_h^{(\tau)}(t) = Q_s^{(\tau)}(t) \;\; \text{vs} \;\; H_1 : Q_h^{(\tau)}(t) \neq Q_s^{(\tau)}(t). \tag{17}$$

Now, suppose that all quantile functions share the same truncated power basis functions $B(t)$ and $Z(t)$. Then, the τ–quantile curves of gene expressions depend on the fixed effects parameter $\boldsymbol{\beta}^{(\tau)}$ and the random effects components $\boldsymbol{U}^{(\tau)}$, and testing the similarity of two gene expression is equivalent to testing the equality of the corresponding fixed effects parameters and random effects components. Further, the predictors of random effects components depend on the estimates of parameters $\boldsymbol{\beta}^{(\tau)T}, \boldsymbol{\alpha}^{(\tau)}, \boldsymbol{\sigma}^{(\tau)}$. Thus, instead of testing the hypothesis (17), one can test the following hypothesis related to the fixed effects parameters and the random effects components.

$$\begin{aligned} H_0 :& (\boldsymbol{\beta}_h^{(\tau)\top}, \boldsymbol{\alpha}_h^{(\tau)}, \boldsymbol{\sigma}_h^{(\tau)}) = (\boldsymbol{\beta}_s^{(\tau)\top}, \boldsymbol{\alpha}_s^{(\tau)}, \boldsymbol{\sigma}_s^{(\tau)}) \;\; \text{vs} \\ H_1 :& (\boldsymbol{\beta}_h^{(\tau)\top}, \boldsymbol{\alpha}_h^{(\tau)}, \boldsymbol{\sigma}_h^{(\tau)}) \neq (\boldsymbol{\beta}_s^{(\tau)\top}, \boldsymbol{\alpha}_s^{(\tau)}, \boldsymbol{\sigma}_s^{(\tau)}). \end{aligned} \tag{18}$$

Based on parameter estimates and its covariance matrix, the following asymptotic statistic for testing the the hypothesis $H_0 : (\boldsymbol{\beta}_h^{(\tau)\top}, \boldsymbol{\alpha}_h^{(\tau)}, \boldsymbol{\sigma}_h^{(\tau)}) = (\boldsymbol{\beta}_s^{(\tau)\top}, \boldsymbol{\alpha}_s^{(\tau)}, \boldsymbol{\sigma}_s^{(\tau)}); (h \neq s)$, is given by

$$\chi_{hs}^{2(\tau)} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_h^{(\tau)} - \hat{\boldsymbol{\beta}}_s^{(\tau)} \\ \hat{\boldsymbol{\alpha}}_h^{(\tau)} - \hat{\boldsymbol{\alpha}}_s^{(\tau)} \\ \hat{\boldsymbol{\sigma}}_h^{(\tau)} - \hat{\boldsymbol{\sigma}}_s^{(\tau)} \end{pmatrix}^\top \boldsymbol{\Xi} \begin{pmatrix} \hat{\boldsymbol{\beta}}_h^{(\tau)} - \hat{\boldsymbol{\beta}}_s^{(\tau)} \\ \hat{\boldsymbol{\alpha}}_h^{(\tau)} - \hat{\boldsymbol{\alpha}}_s^{(\tau)} \\ \hat{\boldsymbol{\sigma}}_h^{(\tau)} - \hat{\boldsymbol{\sigma}}_s^{(\tau)} \end{pmatrix}, \tag{19}$$

where, $\boldsymbol{\Xi} = (\hat{\boldsymbol{V}}^h_{\boldsymbol{\beta}_h^{(\tau)\top}, \boldsymbol{\alpha}_h^{(\tau)}, \boldsymbol{\sigma}_h^{(\tau)}} + \hat{\boldsymbol{V}}^s_{\boldsymbol{\beta}_s^{(\tau)\top}, \boldsymbol{\alpha}_s^{(\tau)}, \boldsymbol{\sigma}_s^{(\tau)}})^{-1}$ and $\hat{\boldsymbol{V}}^h_{\boldsymbol{\beta}_h^{(\tau)\top}, \boldsymbol{\alpha}_h^{(\tau)}, \boldsymbol{\sigma}_h^{(\tau)}}$ and $\hat{\boldsymbol{V}}^s_{\boldsymbol{\beta}_s^{(\tau)T}, \boldsymbol{\alpha}_s^{(\tau)}, \boldsymbol{\sigma}_s^{(\tau)}}$ are the estimated variance–covariance matrices of the estimators $(\boldsymbol{\beta}_h^{(\tau)^\top}, \boldsymbol{\alpha}_h^{(\tau)}, \boldsymbol{\sigma}_h^{(\tau)})$ and $(\boldsymbol{\beta}_s^{(\tau)^\top}, \boldsymbol{\alpha}_s^{(\tau)}, \boldsymbol{\sigma}_s^{(\tau)})$, respectively, which can be computed from (16). When H_0 holds, $\chi_{hs}^{2(\tau)}$ in Equation (19) has asymptotic chi-squared distribution with $(r + q + 1)$ degrees of freedom, where r is known as the dimension number of $\boldsymbol{\beta}^{(\tau)}$ for fixed effects parameters and q is the dimension number of $\boldsymbol{\alpha}^{(\tau)}$ for random effects components.

Furthermore, since the quantile function of gene expression is determined by r-dimensional fixed parameters, q-dimensional random effects components, and scale parameter $\boldsymbol{\sigma}^{(\tau)}$, the pattern of gene expression h can be induced to be a random point in the $(r + q + 1)$ dimensional Euclidean space, which has the asymptotic multivariate normal distribution with mean $(\boldsymbol{\beta}_h^{(\tau)}, \boldsymbol{\alpha}_h^{(\tau)}, \boldsymbol{\sigma}_h^{(\tau)})$ and covariance matrix $\hat{\boldsymbol{V}}^h_{\boldsymbol{\beta}_h^{(\tau)}, \boldsymbol{\alpha}_h^{(\tau)}, \boldsymbol{\sigma}_h^{(\tau)}}$ for $h = 1, \ldots, g$. From this point of view, the statistic $\chi_{hs}^{2(\tau)}$ in (19) is also the Mahalanobis distance between random vectors $(\hat{\boldsymbol{\beta}}_h^{(\tau)}, \boldsymbol{\alpha}_h^{(\tau)}, \boldsymbol{\sigma}_h^{(\tau)})$ and $(\hat{\boldsymbol{\beta}}_s^{(\tau)}, \boldsymbol{\alpha}_s^{(\tau)}, \boldsymbol{\sigma}_s^{(\tau)})$. In terms of the minimum Mahalanobis distance, we can obtain the clustering tree for g gene expressions.

3. Simulation

In this section, we examine the performance of the estimation and chi-square test (19) based on the proposed quantile model for analyzing temporal gene expression data. Extensive simulation studies are carried out to evaluate the accuracy of the estimated quantile curves and the power of the proposed chi-square test (19). The random intercept

QR model and random slope QR model are chosen with five different error terms, in which the error terms are chosen from symmetric distribution (normal), symmetric and heavy tailed distribution (Laplace and Student t) and skewed distribution (skew normal and skew t). Hence, there are ten different scenarios to examine through the simulation study. The steps are given below.

- Thirty-five equally spaced timepoints between [0, 1] are considered for 50 samples. The data were generated using the following mixed model. The true model is assumed as follows:
$$Y_i(t) = f_0(t) + U_i(t) + \epsilon_i(t); \quad t \in [0,1], \tag{20}$$
where, $f_0(t) = \exp\left(\frac{5t}{1+t^3}\right), i = 1, \cdots, 50$.
- Random effects components are assumed to follow a normal distribution with a mean of zero and standard deviation of two, i.e., ($U_i \sim N(0,4)$).
- The following i.i.d errors are considered for true models Equation (20).
  1. Model 1: $\epsilon_i(t) \sim$ Laplace(0,1)
  2. Model 2: $\epsilon_i(t) \sim$ Normal(0,1)
  3. Model 3: $\epsilon_i(t) \sim$ Skew Normal(0,1,1)
  4. Model 4: $\epsilon_i(t) \sim$ Skew t(0,1,1,4)
  5. Model 5: $\epsilon_i(t) \sim$ t(3)

3.1. Model and Parameter Estimation

For model generalization, the following quantile mixed model is considered:

$$Y_i(t) = B^\top(t)\boldsymbol{\beta}^{(\tau)} + Z^\top(t)\boldsymbol{U}_i + \epsilon_i^{(\tau)}, \tag{21}$$

where, $B(t)$ and $Z(t)$ are the basis functions for fixed and random effects components, respectively. The parameter estimation procedure is illustrated below.

- Generate data for $n = 50$ samples using Equation (20) .
- Basis functions for Model (21) are considered as:
  1. Random intercept model: $B(t) = (1, t, t^2, t^3)^\top, \quad Z(t) = (1)$.
  2. Random slope model: $B(t) = (1, t, t^2, t^3)^\top, \quad Z(t) = (1, t)$.
- Forthe random slope model, we consider normal random effects with a diagonal variance–covariance matrix. A Gauss–Hermite quadrature with seven noves is considered to approximate the marginal log-likelihood of Equation (10).
- All parameters are estimated at median ($\tau = 0.50$).
- A total of 500 bootstrap replications are considered for the estimation of the covariance matrix of the estimators $(\hat{\boldsymbol{\beta}}^{(\tau)}, \hat{\boldsymbol{U}}^{(\tau)})$.

Note that, for simplicity, the interior knots are not added in the basis function $B(t)$ in this simulation study. The simulations were also conducted in the scenario for the basis function $B(t)$ with the interior knots and the results are similar to those without the interior knots. Furthermore, the simulations for the other values of quantile τ are omitted.

3.2. Simulation Results for Parameter Estimation

3.2.1. Parameter Estimates of Random Intercept QR Model

Parameter estimates for the 50th quantile ($\tau = 0.50$) of the random intercept QR model are reported below for five different scenarios. The estimated median function, along with the confidence interval of each random intercept QR model, is presented in Figures 1–5. Again, the estimated 50th ($\tau = 0.50$) quantile function is also compared with sample median function in each figure.

- Estimates of parameters in Model 1: $\hat{\boldsymbol{\beta}}^{(\tau)} = [1.19, -5.06, 68.86, -53.29]^\top$, $\hat{\boldsymbol{\alpha}}^{(\tau)} = 1.23$, $\hat{\sigma}^{(\tau)} = 0.56$, $\hat{\Psi}^{(\tau)} = 1.52$.

- Estimates of parameters in Model 2: $\hat{\boldsymbol{\beta}}^{(\tau)} = [1.68, -7.42, 74.74, -56.87]^\top$, $\hat{\boldsymbol{\alpha}}^{(\tau)} = 1.01$, $\hat{\sigma}^{(\tau)} = 0.43$, $\hat{\Psi}^{(\tau)} = 1.02$.
- Estimates of parameters in Model 3: $\hat{\boldsymbol{\beta}}^{(\tau)} = [2.03, -7.46, 74.81, -57.17]^\top$, $\hat{\boldsymbol{\alpha}}^{(\tau)} = 1.14$, $\hat{\sigma}^{(\tau)} = 0.36$, $\hat{\Psi}^{(\tau)} = 1.29$.
- Estimates of parameters in Model 4: $\hat{\boldsymbol{\beta}}^{(\tau)} = [1.88, -7.85, 76.28, -57.92]^\top$, $\hat{\boldsymbol{\alpha}}^{(\tau)} = 1.41$, $\hat{\sigma}^{(\tau)} = 0.49$, $\hat{\Psi}^{(\tau)} = 1.98$.
- Estimates of parameters in Model 5: $\hat{\boldsymbol{\beta}}^{(\tau)} = [1.29, -8.70, 76.78, -57.78]^\top$, $\hat{\boldsymbol{\alpha}}^{(\tau)} = 1.42$, $\hat{\sigma}^{(\tau)} = 0.57$, $\hat{\Psi}^{(\tau)} = 2.01$.

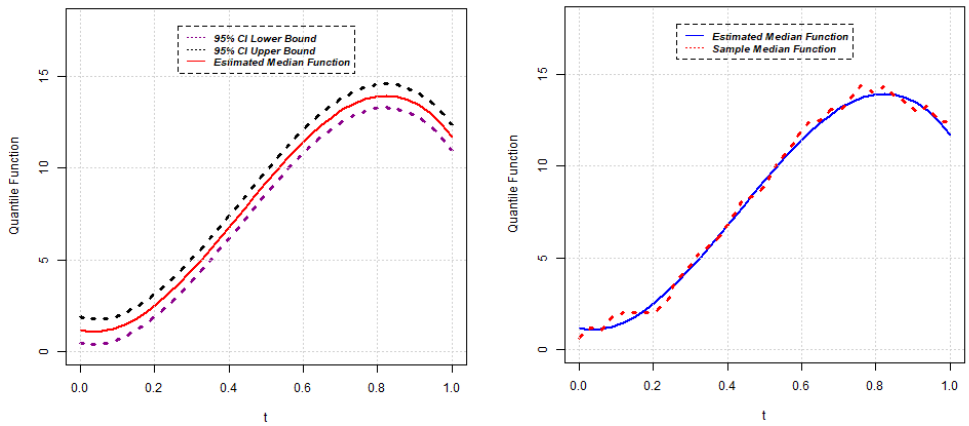

Figure 1. Estimated median function and C.I. (**left**); Estimated median and sample median (**right**) for Model 1.

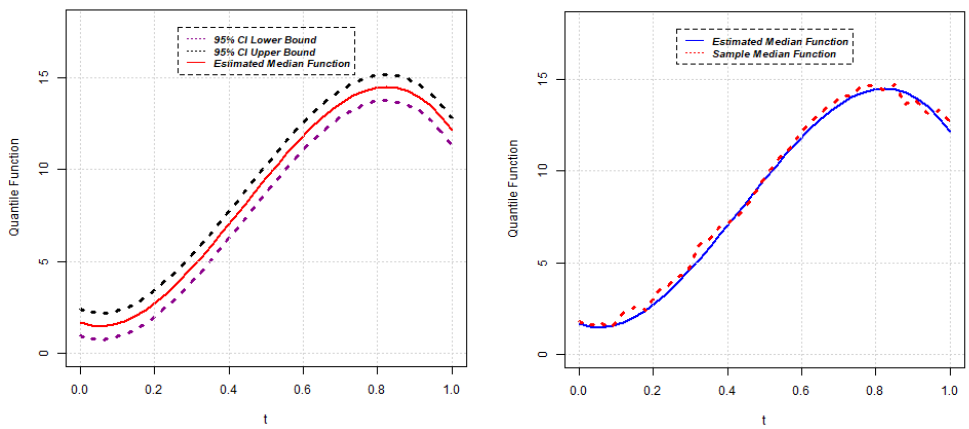

Figure 2. Estimated median function and C.I. (**left**); Estimated median and sample median (**right**) for Model 2.

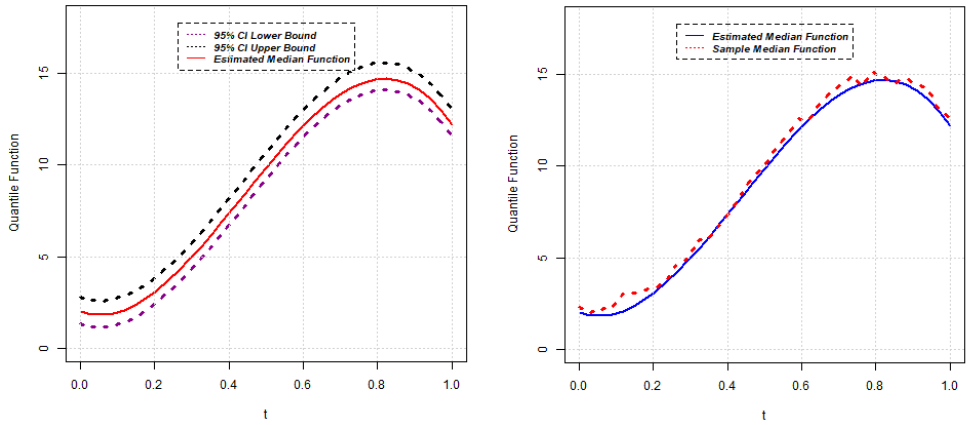

Figure 3. Estimated median function and C.I. (**left**); Estimated median and sample median (**right**) for Model 3.

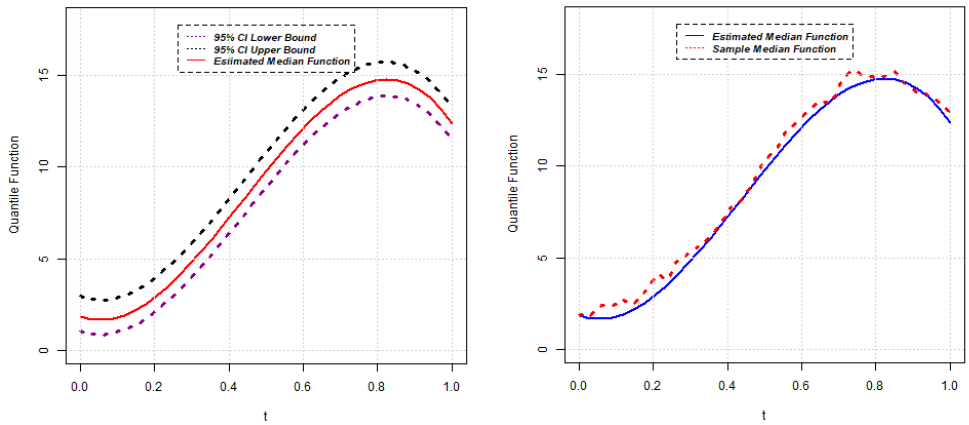

Figure 4. Estimated median function and C.I. (**left**); Estimated median and sample median (**right**) for Model 4.

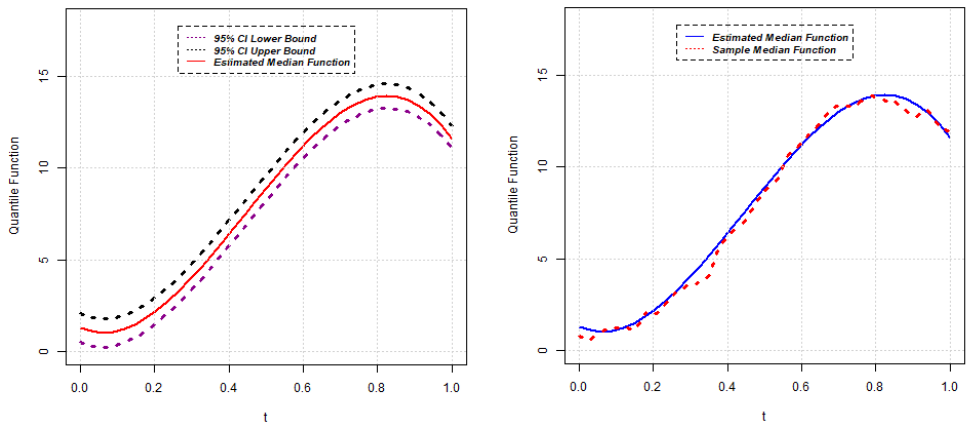

Figure 5. Estimated median function and C.I. (**left**); Estimated median and sample median (**right**) for Model 5.

3.2.2. Parameter Estimates of Random Slope QR Model

Now, the estimates of parameters for the 50th quantile ($\tau = 0.50$) of the random slope QR model are reported below for five different scenarios and estimated median functions; sample median functions, along with confidence intervals, are shown in Figures 6–10.

- Estimates of parameters in Model 1: $\hat{\boldsymbol{\beta}}^{(\tau)} = [1.27, -5.08, 68.86, -53.31]^\top$, $\hat{\boldsymbol{\alpha}}^{(\tau)} = [1.35, 0.67]^\top$, $\hat{\sigma}^{(\tau)} = 0.54$, $\hat{\Psi}^{(\tau)}_{\text{intercept}} = 1.83$, $\hat{\Psi}^{(\tau)}_{\text{time}} = 0.45$.

- Estimates of parameters in Model 2: $\hat{\boldsymbol{\beta}}^{(\tau)} = [0.98, -6.43, 72.80, -55.64]^\top$, $\hat{\boldsymbol{\alpha}}^{(\tau)} = [1.29, 0.68]^\top$, $\hat{\sigma}^{(\tau)} = 0.41$, $\hat{\Psi}^{(\tau)}_{\text{intercept}} = 1.68$, $\hat{\Psi}^{(\tau)}_{\text{time}} = 0.46$.
- Estimates of parameters in Model 3: $\hat{\boldsymbol{\beta}}^{(\tau)} = [1.92, -7.43, 74.85, -57.15]^\top$, $\hat{\boldsymbol{\alpha}}^{(\tau)} = [1.79, 0.73]^\top$, $\hat{\sigma}^{(\tau)} = 0.35$, $\hat{\Psi}^{(\tau)}_{\text{intercept}} = 3.20$, $\hat{\Psi}^{(\tau)}_{\text{time}} = 0.53$.
- Estimates of parameters in Model 4: $\hat{\boldsymbol{\beta}}^{(\tau)} = [2.30, -7.83, 76.31, -57.87]^\top$, $\hat{\boldsymbol{\alpha}}^{(\tau)} = [1.67, 1.06]^\top$, $\hat{\sigma}^{(\tau)} = 0.46$, $\hat{\Psi}^{(\tau)}_{\text{intercept}} = 2.79$, $\hat{\Psi}^{(\tau)}_{\text{time}} = 1.13$.
- Estimates of parameters in Model 5: $\hat{\boldsymbol{\beta}}^{(\tau)} = [1.25, -8.53, 76.79, -57.92]^\top$, $\hat{\boldsymbol{\alpha}}^{(\tau)} = [1.47, 0.53]^\top$, $\hat{\sigma}^{(\tau)} = 0.57$, $\hat{\Psi}^{(\tau)}_{\text{intercept}} = 2.16$, $\hat{\Psi}^{2}_{\text{time}} = 0.29$.

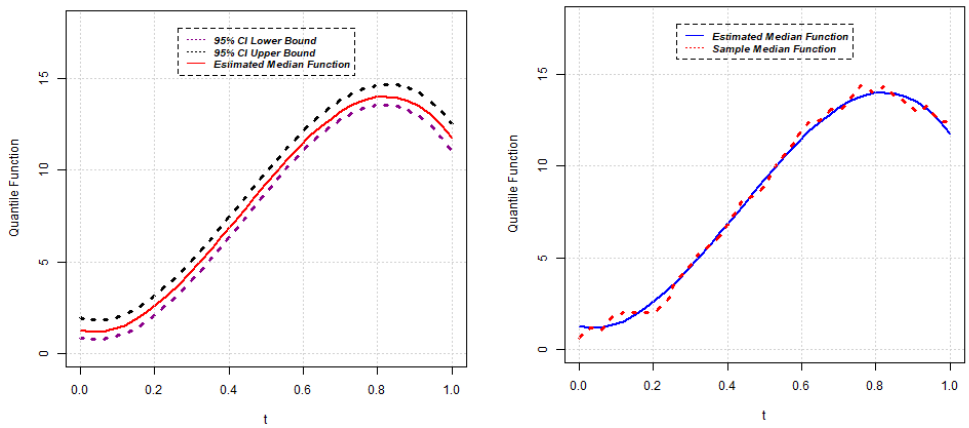

Figure 6. Estimated median function and C.I. (**left**); Estimated median and sample median (**right**) for Model 1.

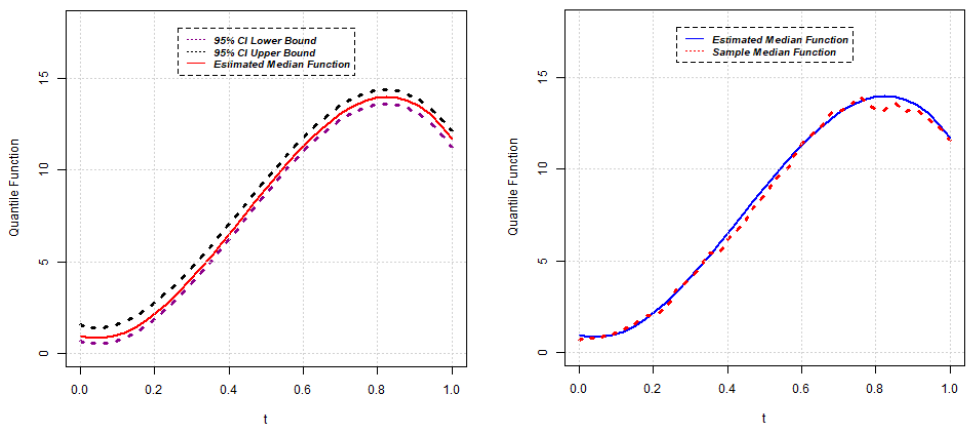

Figure 7. Estimated median function and C.I. (**left**); Estimated median and sample median (**right**) for Model 2.

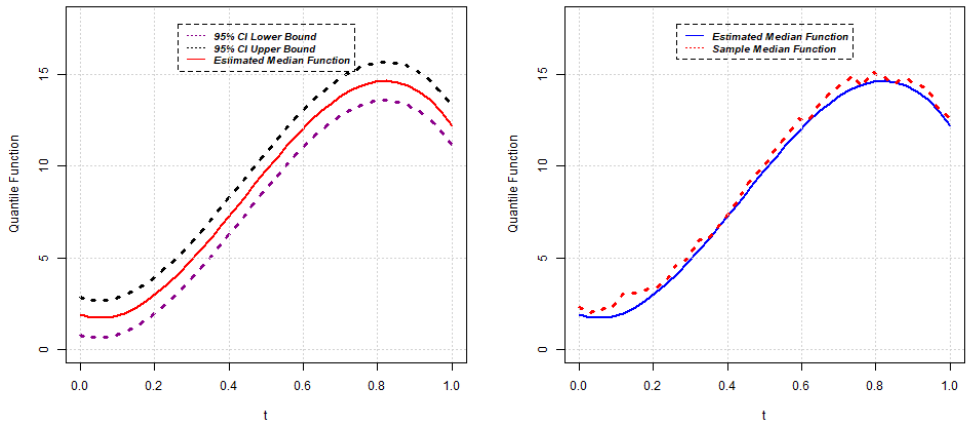

Figure 8. Estimated median function and C.I. (**left**); Estimated median and sample median (**right**) for Model 3.

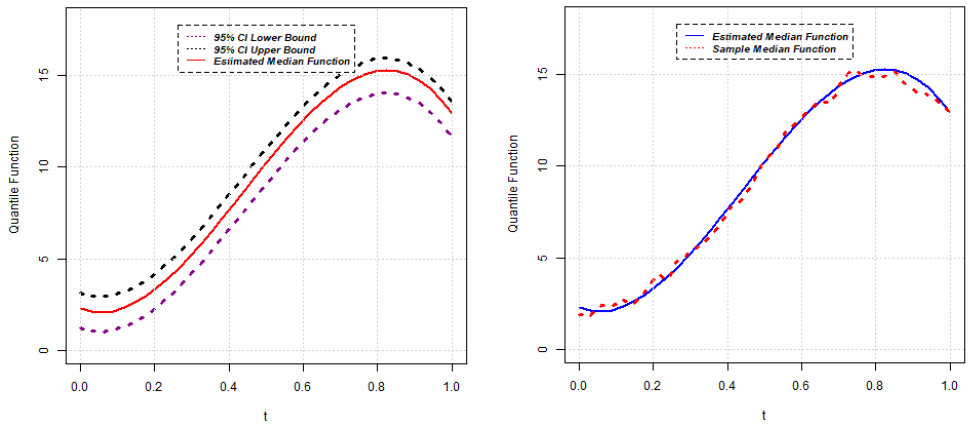

Figure 9. Estimated median function and C.I. (**left**); Estimated median and sample median (**right**) for Model 4.

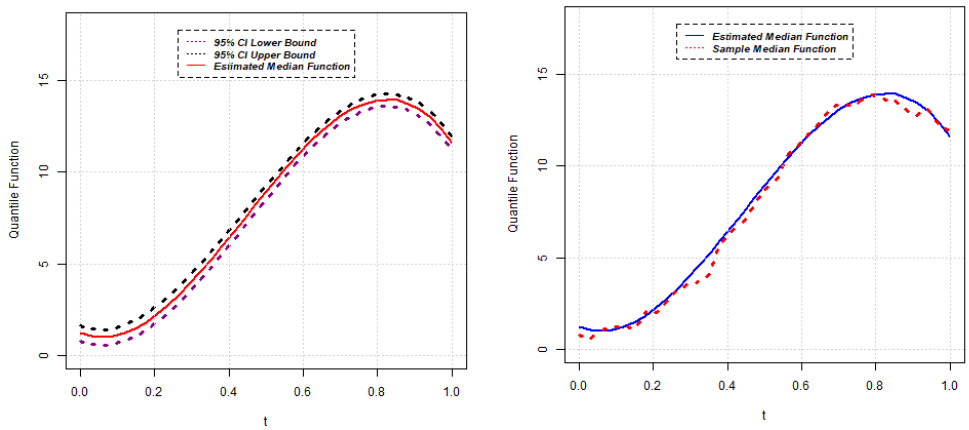

Figure 10. Estimated median function and C.I. (**left**); Estimated median and sample median (**right**) for Model 5.

3.3. Power Analysis of Proposed Test Statistic

In this section, the power of the chi-square test presented in Equation (19) is evaluated by considering the alternative model. A constant number, m is added to this alternative model.

The alternative model considered in simulation is as follows:

$$[Y_i(t)]_{\text{new}} = \exp(\frac{5t}{1+t^3}) + U_i(t) + \epsilon_i(t) + m \tag{22}$$

where, $m \in \{0, 0.50, 1.00, 1.50, 2.00, 2.50, 3.00\}$. The L_2 distance between true model and alternative model is m. Here, we assume that random effects (U_i) and error (ϵ_i) follow the same distribution. The power analysis is carried out for both the random slope QR model and random intercept QR model. All simulation results are performed for a sample size of $n = 30, 50$ and 75 at a level of significance of $\alpha = 0.05$ and quantile value of $\tau = 0.25, 0.50, 0.75$ with 1000 replications. Tables 1–3 report empirical powers for five random intercept QR models and five random slope models with the quantiles $\tau = 0.25, 0.5$ and 0.75. Power analysis results demonstrate that the statistical performance of the proposed chi-square test statistic is good for both random intercept models and random slope models at a 5% level of significance. This indicates that this test statistic can be applied to determine the similarity of two gene expressions.

Table 1. Empirical Powers of $\chi^{2(\tau)}_{hs}$ for intercept models and slope models with $\tau = 0.25$.

		(a) Intercept Models			**(b) Slope Models**		
Models	**Distance**	$n = 30$	$n = 50$	$n = 75$	$n = 30$	$n = 50$	$n = 75$
Model 1	0.00	0.055	0.052	0.051	0.042	0.049	0.052
	0.50	0.077	0.083	0.073	0.066	0.066	0.054
	1.00	0.144	0.175	0.174	0.122	0.151	0.197
	1.50	0.293	0.347	0.397	0.234	0.320	0.408
	2.00	0.476	0.576	0.628	0.420	0.551	0.659
	2.50	0.693	0.777	0.828	0.639	0.752	0.860
	3.00	0.837	0.900	0.954	0.808	0.873	0.943
Model 2	0.00	0.055	0.047	0.049	0.044	0.051	0.047
	0.50	0.094	0.065	0.077	0.074	0.063	0.078
	1.00	0.163	0.171	0.215	0.122	0.155	0.208
	1.50	0.325	0.365	0.453	0.242	0.315	0.443
	2.00	0.533	0.597	0.709	0.430	0.556	0.686
	2.50	0.728	0.811	0.868	0.653	0.774	0.851
	3.00	0.871	0.922	0.958	0.826	0.899	0.944
Model 3	0.00	0.068	0.057	0.051	0.059	0.050	0.046
	0.50	0.108	0.082	0.086	0.078	0.065	0.080
	1.00	0.194	0.204	0.218	0.138	0.153	0.178
	1.50	0.349	0.417	0.494	0.255	0.346	0.403
	2.00	0.563	0.668	0.739	0.449	0.601	0.661
	2.50	0.756	0.847	0.897	0.653	0.797	0.842
	3.00	0.885	0.957	0.964	0.831	0.920	0.950
Model 4	0.00	0.066	0.058	0.053	0.057	0.046	0.054
	0.50	0.086	0.084	0.086	0.072	0.057	0.061
	1.00	0.159	0.172	0.194	0.133	0.148	0.173
	1.50	0.298	0.375	0.428	0.252	0.326	0.389
	2.00	0.523	0.621	0.697	0.423	0.572	0.662
	2.50	0.717	0.816	0.865	0.669	0.774	0.842
	3.00	0.871	0.933	0.952	0.818	0.901	0.942
Model 5	0.00	0.056	0.053	0.049	0.057	0.043	0.046
	0.50	0.085	0.082	0.085	0.085	0.062	0.071
	1.00	0.144	0.167	0.187	0.149	0.157	0.195
	1.50	0.270	0.336	0.402	0.253	0.323	0.422
	2.00	0.460	0.558	0.671	0.448	0.553	0.675
	2.50	0.651	0.762	0.844	0.634	0.748	0.851
	3.00	0.806	0.890	0.947	0.786	0.889	0.949

Table 2. Empirical Powers of $\chi_{hs}^{2(\tau)}$ for intercept models and slope models with $\tau = 0.50$.

		(a) Intercept Models			(b) Slope Models		
Models	**Distance**	$n = 30$	$n = 50$	$n = 75$	$n = 30$	$n = 50$	$n = 75$
	0.00	0.071	0.047	0.051	0.051	0.057	0.053
	0.50	0.095	0.091	0.086	0.076	0.087	0.072
	1.00	0.193	0.240	0.270	0.154	0.182	0.227
Model 1	1.50	0.360	0.489	0.574	0.313	0.413	0.505
	2.00	0.597	0.751	0.812	0.521	0.664	0.774
	2.50	0.807	0.903	0.940	0.720	0.849	0.913
	3.00	0.919	0.976	0.991	0.874	0.951	0.981
	0.00	0.066	0.049	0.052	0.048	0.050	0.046
	0.50	0.092	0.089	0.100	0.080	0.069	0.059
	1.00	0.190	0.239	0.318	0.158	0.176	0.227
Model 2	1.50	0.396	0.479	0.635	0.317	0.397	0.491
	2.00	0.638	0.730	0.854	0.536	0.664	0.761
	2.50	0.849	0.922	0.967	0.749	0.857	0.927
	3.00	0.944	0.977	0.990	0.876	0.952	0.978
	0.00	0.078	0.047	0.048	0.065	0.051	0.057
	0.50	0.102	0.105	0.096	0.087	0.063	0.073
	1.00	0.211	0.240	0.320	0.152	0.196	0.226
Model 3	1.50	0.383	0.511	0.615	0.310	0.414	0.488
	2.00	0.628	0.770	0.842	0.493	0.665	0.752
	2.50	0.820	0.924	0.955	0.711	0.850	0.911
	3.00	0.937	0.990	0.992	0.872	0.954	0.976
	0.00	0.076	0.057	0.051	0.044	0.045	0.047
	0.50	0.101	0.080	0.102	0.084	0.070	0.071
	1.00	0.202	0.223	0.291	0.151	0.191	0.226
Model 4	1.50	0.379	0.510	0.606	0.328	0.436	0.508
	2.00	0.626	0.754	0.847	0.559	0.679	0.767
	2.50	0.835	0.911	0.952	0.755	0.855	0.909
	3.00	0.934	0.977	0.985	0.881	0.951	0.975
	0.00	0.058	0.055	0.052	0.048	0.056	0.048
	0.50	0.092	0.084	0.089	0.079	0.074	0.080
	1.00	0.183	0.222	0.273	0.157	0.166	0.246
Model 5	1.50	0.374	0.491	0.587	0.316	0.420	0.513
	2.00	0.608	0.724	0.836	0.522	0.672	0.786
	2.50	0.796	0.897	0.942	0.741	0.851	0.924
	3.00	0.922	0.966	0.983	0.874	0.936	0.982

Table 3. Empirical Powers of $\chi_{hs}^{2(\tau)}$ for intercept models and slope models with $\tau = 0.75$.

		(a) Intercept Models			(b) Slope Models		
Models	**Distance**	$n = 30$	$n = 50$	$n = 75$	$n = 30$	$n = 50$	$n = 75$
	0.00	0.073	0.056	0.047	0.069	0.058	0.044
	0.50	0.098	0.078	0.069	0.081	0.071	0.071
	1.00	0.163	0.170	0.197	0.150	0.158	0.177
Model 1	1.50	0.303	0.359	0.415	0.285	0.320	0.396
	2.00	0.499	0.572	0.661	0.447	0.551	0.666
	2.50	0.686	0.777	0.852	0.639	0.758	0.856
	3.00	0.818	0.903	0.955	0.795	0.894	0.943
	0.00	0.058	0.050	0.050	0.050	0.057	0.048
	0.50	0.083	0.066	0.082	0.064	0.070	0.074
	1.00	0.164	0.170	0.198	0.139	0.156	0.199
Model 2	1.50	0.332	0.380	0.455	0.251	0.346	0.442
	2.00	0.535	0.634	0.724	0.449	0.567	0.696
	2.50	0.740	0.816	0.881	0.655	0.771	0.863
	3.00	0.880	0.923	0.960	0.809	0.899	0.941

Table 3. *Cont.*

Models	Distance	(a) Intercept Models			(b) Slope Models		
		$n = 30$	$n = 50$	$n = 75$	$n = 30$	$n = 50$	$n = 75$
Model 3	0.00	0.075	0.046	0.054	0.077	0.046	0.050
	0.50	0.101	0.086	0.084	0.098	0.078	0.081
	1.00	0.173	0.202	0.226	0.178	0.169	0.214
	1.50	0.347	0.440	0.486	0.303	0.371	0.445
	2.00	0.539	0.677	0.757	0.463	0.609	0.690
	2.50	0.739	0.864	0.910	0.663	0.817	0.855
	3.00	0.873	0.940	0.958	0.817	0.921	0.951
Model 4	0.00	0.066	0.050	0.051	0.055	0.044	0.044
	0.50	0.079	0.081	0.082	0.077	0.055	0.072
	1.00	0.166	0.188	0.203	0.132	0.163	0.205
	1.50	0.319	0.388	0.451	0.273	0.360	0.446
	2.00	0.523	0.638	0.680	0.468	0.610	0.726
	2.50	0.717	0.819	0.873	0.679	0.803	0.887
	3.00	0.868	0.946	0.957	0.831	0.925	0.966
Model 5	0.00	0.046	0.046	0.054	0.056	0.051	0.047
	0.50	0.070	0.066	0.081	0.070	0.073	0.076
	1.00	0.131	0.154	0.187	0.136	0.143	0.177
	1.50	0.264	0.333	0.400	0.281	0.320	0.413
	2.00	0.456	0.567	0.644	0.449	0.552	0.688
	2.50	0.664	0.773	0.831	0.633	0.757	0.867
	3.00	0.833	0.913	0.940	0.817	0.881	0.948

4. Application: Gene Expression Data

In this section, the linear quantile mixed model given in Section 2 is applied to analyze a real dataset of 18 gene expressions in *P. aeruginosa* expressed in 24 conditions. Some descriptive measures are discussed at the beginning of this section. Towards the end of this section, the similarities between genes are investigated using the proposed test statistic (Equation (19)) based on quantile functions.

4.1. Data

A description of 18 genes is reported in Table 4. The data observed for these genes are first analyzed by Fang et al. [1] and then by Deng et al. [10] Initially, polymerase chain reaction (PCR) was performed to amplify regions of *P. aeruginosa* virulence factors. The primers were synthesized using PAO1 genome data according to Duan [18]. PAO1 chromosomal DNA was used as a PCR template and amplified promoter regions were cloned into XhoI-BamHI restriction sites of the plasmid PMS402. Then plasmids were inserted into PAO1 using electroporation. More details regarding DNA manipulation, PCR and transformation procedures can be found in Duan [18]. The promoter activity was measured as counts per second (CPS) of light production using a Victor2 Multilabel counter. TSBDC minimal medium containing EDTA (400 µg/mL) and (50 µg/mL) $Fecl_3$ were used to assay gene expression. The reporter strains were grown overnight and the resulting culture was diluted into 1:200 proportion in a 96-well microtiter plate. Then, the promoter activity of the virulence factors was measured under 24 biological conditions every 30 min for 21 h. These genes were considered as quorum-sensing or quorum-sensing-regulated genes, which play an important role in bio-films formation. Thus, this dataset consists of 18 genes in *P. aeruginosa* (Table 4) and each gene was observed at 43 consecutive timepoints under 24 conditions. Therefore, the data for each gene consist of 1032 observations and, in total, 18,576 observations are measured for 18 genes.

Table 4. 18 genes in *P. aeruginosa* expression.

Code	Name	Protein	Ratio	Remarks
A6	PA5283	Probable Trancriptional regulator	99.68%	48% similar to putative transcriptional regulator (Bacilus subtilis)
B3	PA2975 (rluc)	Ribosomal large subunit	99.68%	Transcription, RNA processing and degradation
B4	PA4991	Hypothetical Protein	100%	Unknown
B5	PA5237	Conserved hypothetical Protein	100%	87% similar to hypothetical yigC gene product of E. coli
C4	PA0287 (gpuP)	3-guanidinopropionate transport protein	100%	Transport of small molecules
D1	PA3115 (fimV)	Motility protein FimV	100%	Membrane proteins; Motility and Attachment
D2	PA3879 (narL)	Two-component response regulator NarL	99.67%	74% similar to E. coli NarL protein
D3	PA0894	Hypothetical Protein	99.02%	Unknown
E5	PA1875	Probable outer membrane protein precursor	100%	41% similar to alkaline protease secretion protein AprF
E6	PA0573	Hypothetical Protein	100%	Unknown
F2	PA3902	Hypothetical Protein	100%	Unknown
F3	PA3212	Probable ATP-binding component of ABC transporter	100%	65% similar to putative amino acid abc transporter, ATP-binding Protein (Helicobacter Pylori J99)
F5	PA2997 (nqrC)	Na+translocating NADH: ubiquinone oxidoreductase subunit Nrq3	100%	Energy metabolism
G2	PA0649 (trpG)	Anthranilate synthase component II	100%	Energy metabolism; Biosynthesis of cofactors, prosthetic groups and carriers; Amino acid biosynthesis and metabolism
G5	PA1748	Probable enoyl-CoA hydratase/isomerase	98.20%	61% similar to putative enoyl-coA hydratase EchA3 (Mycobacterium tuberculosis)
G6	PA3771	Probable trancriptional regulator	99.22%	54% similar to a region of putative regulatory protein (Streptomyces coelicolor)
H3	PA1841	Hypothetical protein	100%	43% similar to hypothetical yeaK gene product of (E. coli)
S70	$\sigma 70$	σ factor		As a control

4.2. Exploratory Analysis

In this section, a summary is provided of the selected gene expression data. Gene expressions, $g(t)$ are obtained on a log scale under 24 conditions. Figure 11 exhibits gene expression curves of genes PA2975(rluc) and PA0573, respectively. From this figure, it is evident that the expressions for genes PA2975(rluc) and PA0573 are very different under different biological conditions. Figure 12 shows a histogram of expressions on PA2975(rluc) and PA0573, respectively. This figure reveals that teh data are severely skewed (negative) and show non-normality signs. Figures 13 and 14 present the distributions of expressions for PA2975(rluc) and PA0573 with respect to time and conditions, respectively. Both boxplots demonstrate that heterogeneity is present in both PA2975(rluc)(B3) and PA0573(E6) gene expressions. More importantly, both parts of the figure show that the gene expression dataset consists of outliers. Hence, Figures 12–14 suggest that the normality assumption for linear mixed model may be inappropriate for analyzing gene expression data. For this reason, the quantile models are going to be used to determine gene trajectory in terms of different quantile values. This method might provide more significant and insightful results than the mean regression model.

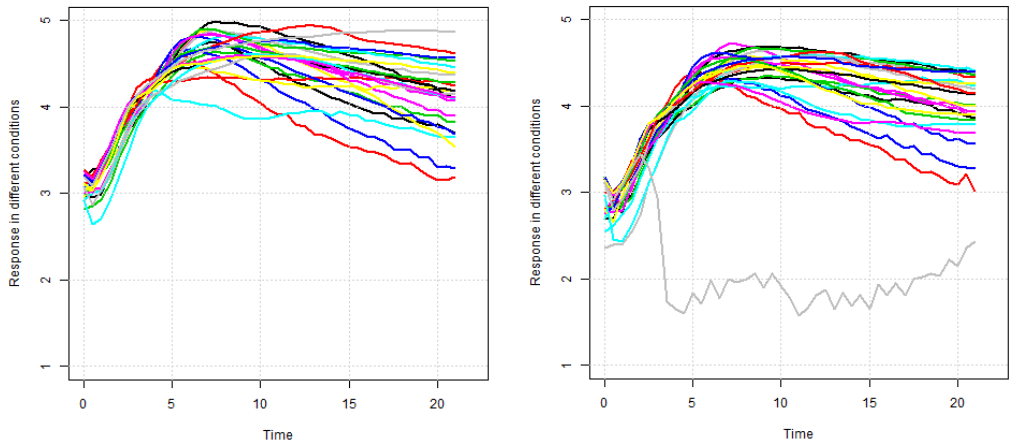

Figure 11. Gene expressions of PA2975(rluc)(B3) (**left**) and PA0573(E6) (**right**).

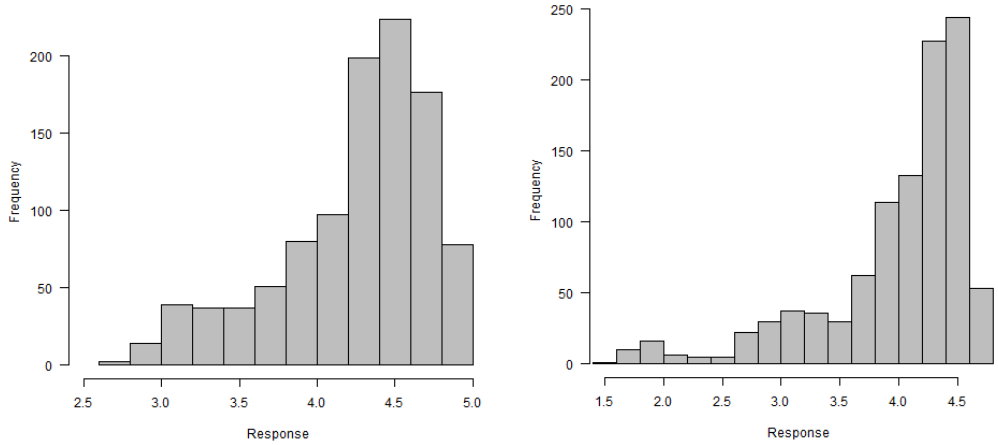

Figure 12. Histograms of gene expressions of PA2975(rluc)(B3) (**left**) and PA0573(E6) (**right**).

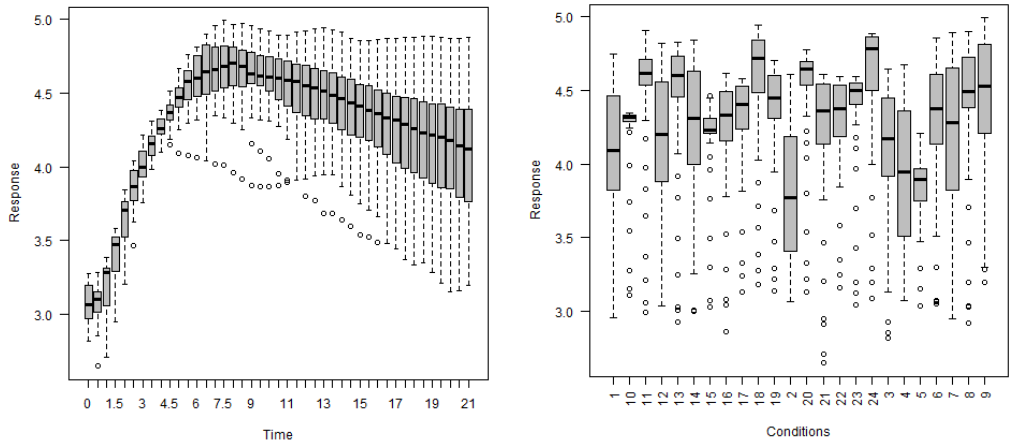

Figure 13. Box-plots of gene expressions PA2975(rluc)(B3) regarding time (**left**) and conditions (**right**).

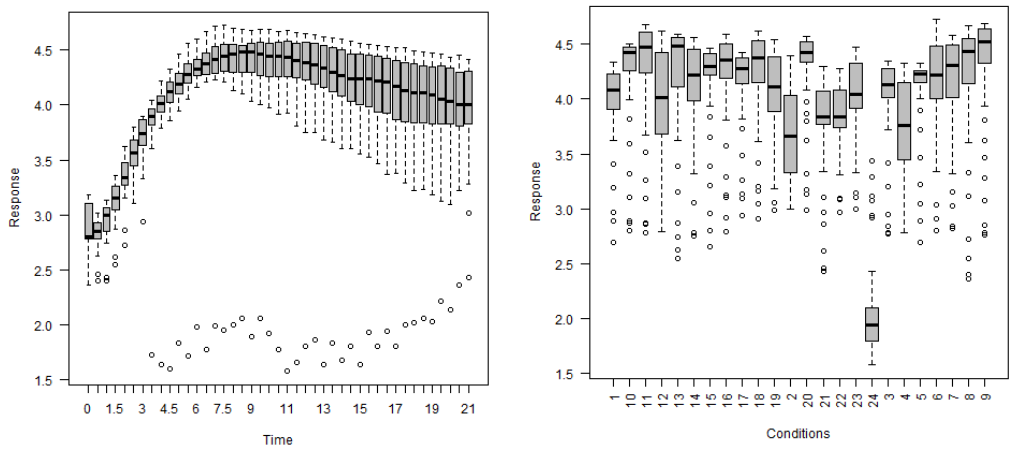

Figure 14. Box-plots of gene expressions PA0573(E6) regarding time (**left**) and conditions (**right**).

4.3. Model and Parameter Estimates

Let $Y_i(t)$ be the measurements for condition i at time t. The linear quantile mixed model is considered for estimation purposes. Since no additional covariates are available, polynomial basis functions are considered for fixed effects in this analysis.

$$Y_i(t) = [B(t)]^\top \beta^{(\tau)} + [Z(t)]^\top U^{(\tau)} + \epsilon_i^{(\tau)}(t); \;\; (i = 1, \cdots, 24) \;\; (t = 0, 0.5, \cdots, 21). \quad (23)$$

Random Intercept Model: $B(t) = (1, t, t^2, t^3, t^4)^\top$ and $Z(t) = (1)$.
Random Slope Model: $B(t) = (1, t, t^2, t^3, t^4)^\top$ and $Z(t) = (1, t)^\top$.

For each gene expression, the same fixed effects and random effects structures are considered when estimating the parameters in both the random intercept model and random slope model. However, AIC is found to be lower for the random intercept model than for the random slope model. Thus, the random intercept model is used for further analysis, instead of the random slope model. Table 5 reports the values of estimated parameters for quantile functions with quantile $\tau = 0.25, 0.50, 0.75$ for PA2975(rluc)(B3) and PA0573(E6). Figures 15 and 16 show the estimated quantile functions and confidence intervals for genes PA2975(rluc)(B3) and PA0573(E6), respectively. Figures 17 and 18 present the estimated quantile functions, along with the sample quantile functions, for gene PA2975(rluc)(B3) and PA0573(E6). Moreover, Figure 19 shows the estimated median functions for 18 genes in *P. aeruginosa*.

Table 5. Parameter estimates of quantile model for genes PA2975(rluc)(B3)and PA0573(E6).

	PA2975(rluc)(B3)			**PA0573(E6)**		
Estimate	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
β_0	2.7592	2.843	2.925	2.569158	2.625809	2.785331
β_1	0.54199	0.53962	0.53468	0.465631	0.494351	0.4591522
β_2	-0.054218	-0.054119	-0.052799	-0.039942	-0.042743	-0.038572
β_3	0.002192	0.0021183	0.0020676	0.001215	0.001262	0.001171
β_4	-0.000031	-0.000029	-0.000029	-0.000011	-0.000008	-0.000009
$\hat{\sigma}$	0.068	0.096	0.072	0.084	0.101	0.076
AIC	43.43	143.56	166.91	466.15	241.57	252.46

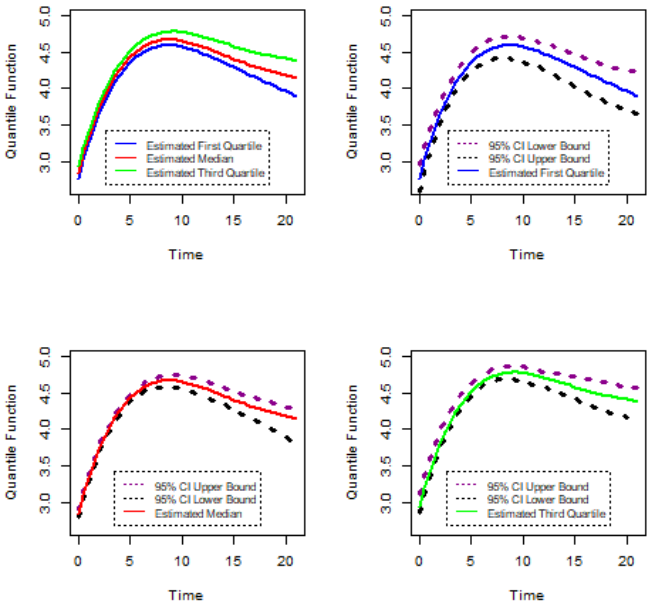

Figure 15. Confidence intervals of gene PA2975(rluc)(B3).

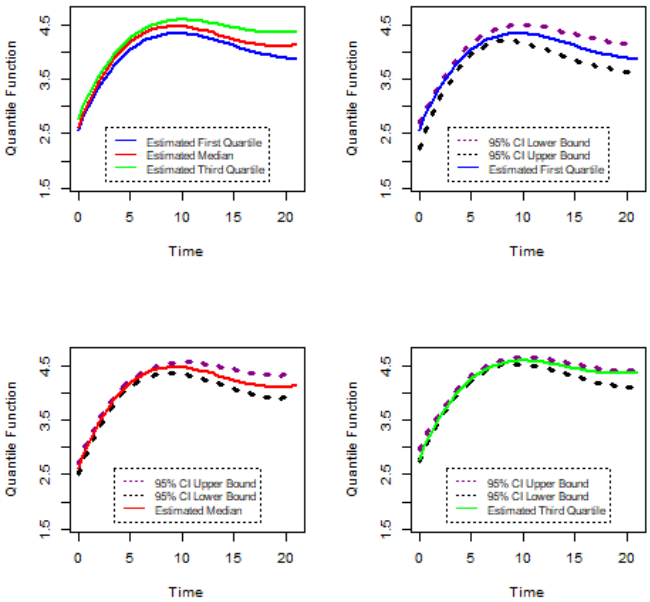

Figure 16. Confidence intervals of gene and PA0573(E6).

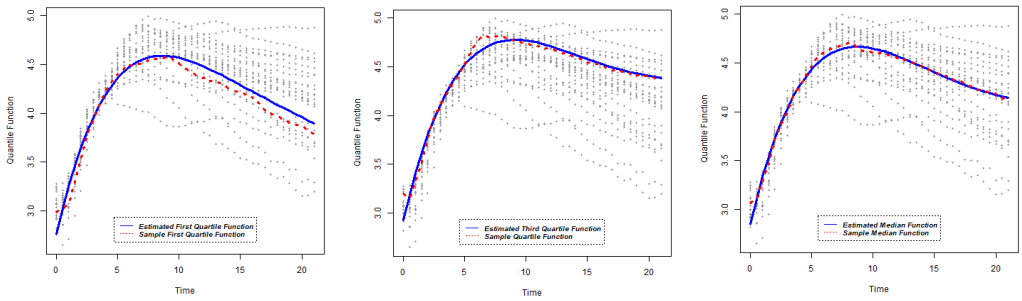

Figure 17. Quantile functions and sample quantile functions of PA2975(rluc)(B3): lower quartile (**left**); median (**center**); upper quartile (**right**).

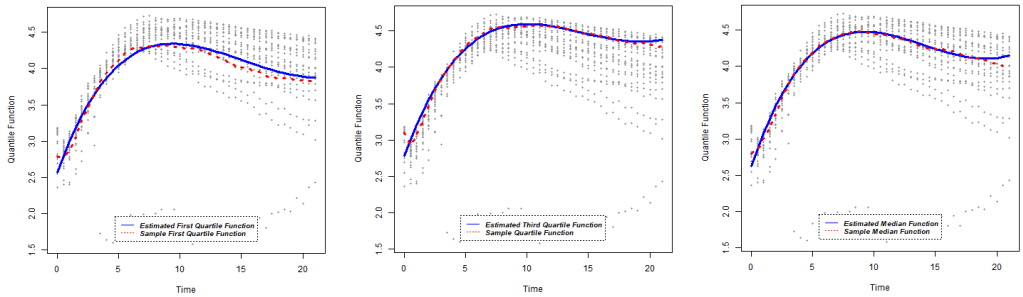

Figure 18. Quantile functions and sample quantile functions of PA0573(E6): lower quartile (**left**); median (**center**); upper quartile (**right**).

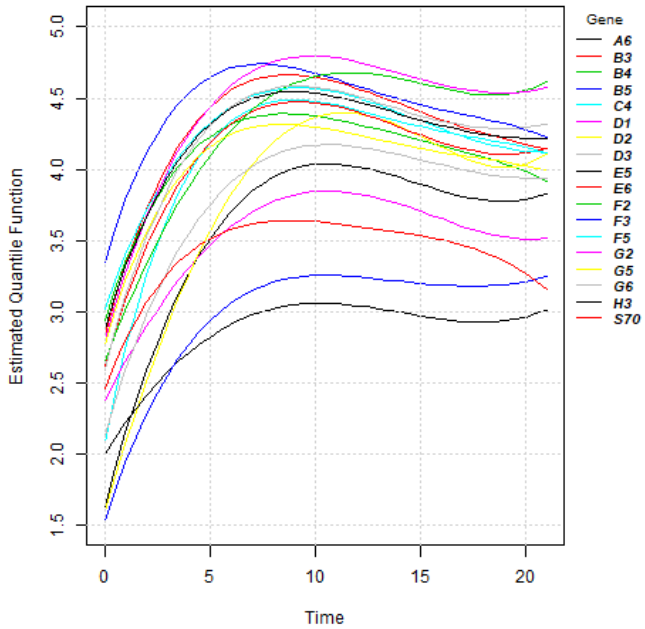

Figure 19. Estimated median functions of 18 genes in *P. aeruginosa*.

4.4. Test of Gene Similarity

To find the similarity between genes, pairwise comparisons of all the gene expressions of 18 genes in *P. aeruginosa* were investigated in terms of their quantile functions, since gene expression curves depend on fixed effects parameters and random effects components. For $\tau = 0.50$ the following hypotheses were tested.

$$\begin{aligned} H_0 &: (\boldsymbol{\beta}_h^{(\tau)}, \boldsymbol{u}_h^{(\tau)}) = (\boldsymbol{\beta}_s^{(\tau)}, \boldsymbol{u}_s^{(\tau)}) \quad \text{vs} \\ H_1 &: (\boldsymbol{\beta}_h^{(\tau)}, \boldsymbol{u}_h^{(\tau)}) \neq (\boldsymbol{\beta}_s^{(\tau)}, \boldsymbol{u}_s^{(\tau)}) \end{aligned} \tag{24}$$

where $\boldsymbol{\beta}_h^{(\tau)}$ and $\boldsymbol{\beta}_s^{(\tau)}$ are the set of parameters $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ for hth gene and sth gene. When H_0 holds, $\chi_{hs}^{2(\tau)}$, Equation (19) has an asymptotic chi-squared distribution with $(r+1)$ degrees of freedom, where r is known as the basis of the regression model. Table 6 reports the Mahalanobis distance between 18 genes using the results of parameter estimates and chi-square statistic for $\tau = 0.50$. Moreover, considering a significance level of $\alpha = 0.05$, Table 7 reports gene similarity. Figure 20 presents clustering tree of 18 genes in *P. aeruginosa*. Based on Figure 20 and Table 7, it can be said that PA1841(H3) shows no significant difference to PA2975(rluc)(B3), PA4991(B4), PA0573(E6), PA2997(F5) and PA1748(G5). In addition, the quantile function of PA3771(G6) shows no difference to the quantile function of PA0287(C4), PA0573(E6), PA0649(G2). Furthermore, PA3212(F3) shows similarities with PA5283(A6) and PA1875(E5).

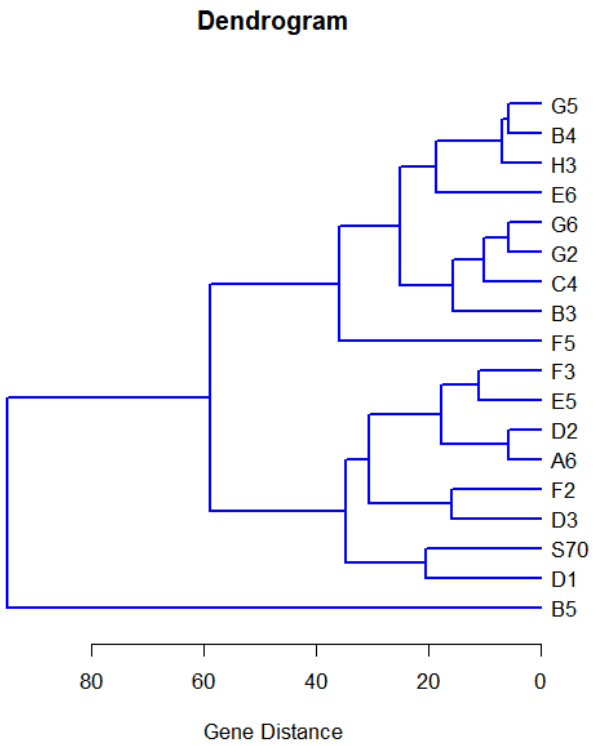

Figure 20. The clustering tree of 18 genes in *P. aeruginosa*.

Table 6. Mahalanobis distance between genes.

Gene	A6	B3	B4	B5	C4	D1	D2	D3	E5	E6	F2	F3	F5	G2	G5	G6	H3
B3	222.58																
B4	145.73	17.46															
B5	331.40	53.15	62.8														
C4	41.58	21.14	20.96	54.61													
D1	29.88	127.88	64.77	176.25	66.30												
D2	5.78	77.32	79.50	105.09	29.10	29.56											
D3	12.93	62.89	68.93	145.2	22.02	29.05	22.91										
E5	18.15	35.27	41.42	49.32	27.38	18.92	24.82	26.35									
E6	159.01	39.13	34.87	190.01	14.69	78.55	51.99	28.54	20.33								
F2	35.29	33.85	34.91	54.04	29.13	33.34	49.49	15.86	49.9	17.52							
F3	9.90	43.38	66.94	92.28	11.63	25.66	18.09	11.51	11.01	24.65	36.72						
F5	120.18	39.42	19.50	37.26	31.58	71.71	93.65	88.4	52.68	55.69	43.79	74.78					
G2	185.55	15.76	53.09	54.37	11.47	105.15	53.06	39.77	27.35	31.29	19.38	25.87	51.5				
G5	104.23	25.73	5.68	62.91	11.67	71.79	61.42	40.07	28.85	11.2	36.9	39.48	25.01	52.04			
G6	62.10	10.25	28.59	36.92	8.99	82.18	35.43	35.28	16.29	13.19	33.96	19.79	41.33	5.85	22.08		
H3	104.05	8.72	7.82	35.07	21.29	78.07	71.37	45.62	36.34	9.94	25.01	39.33	24.02	25.64	6.22	17.28	
σ70	55.17	59.36	29.27	75.08	55.16	20.55	41.38	34.36	23.2	51.65	56.12	39.25	51.18	82.15	28.1	59.18	40.18

Table 7. Gene similarity.

Gene	A6	A6	A6	B3	B3	B4	B4	C4	C4
Gene	D2	D3	F3	G6	H3	G5	H3	F3	G2
Distance	5.78	12.93	9.90	10.25	8.72	5.68	7.82	11.63	11.47
Gene	C4	C4	D3	E5	E6	E6	E6	G2	G5
Gene	G5	G6	F3	F3	G5	G6	H3	G6	H3
Distance	11.67	8.99	11.51	11.01	11.20	13.19	9.94	5.85	6.22

5. Concluding Remarks

Gene expression analysis usually tracks the expression values of a large number of genes simultaneously under different biological conditions, and the rate of change may not be consistent at different quantiles. This research uses a linear quantile mixed effect model to analyze the gene expression trajectory. Gene expression models are constructed using basis functions and a statistical test is proposed to determine the similarity of the

genes based on estimated fixed effects parameters and random effects components from the linear quantile mixed model. In simulation studies, both the random intercept and random slope models are considered to examine the performance of proposed estimation and test statistics. The simulation results indicate that the chi-square test statistic performs well in all circumstances to test the similarity of quantile functions. Finally, as an illustration, a linear quantile mixed model is applied to a real dataset of 18 genes in *P. aeruginosa* expressed in 24 conditions, and the similarity between genes is investigated using the proposed test statistic based on quantile functions. Moreover, a clustering tree is also shown using a complete linkage method. This research suggests that this proposed test statistic may help bio-medical scientists test the similarity between genes in terms of their quantile functions. In a simulation study, instead of true quantile functions, estimated quantile functions are compared with sample quantile functions, which is a limitation of this research. This research can be further extended by considering a changeable variance and covariance structure for the gene expression data.

Author Contributions: Conceptualization, D.D. and M.H.C.; methodology, D.D. and M.H.C.; software, D.D. and M.H.C.; validation, D.D. and M.H.C.; formal analysis, D.D. and M.H.C.; investigation, D.D. and M.H.C.; resources, D.D.; data curation, D.D. and M.H.C.; writing—original draft preparation, D.D. and M.H.C.; writing—review and editing, D.D. and M.H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from other researcher and are available with the permission of this researcher.

Acknowledgments: The authors are very grateful to the editor, associate editor and three referees for their careful reading and valuable comments, which have greatly improved this paper. The first author of this work is partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fang, H.B.; Deng, D.; Tian, G.L.; Shen, L.; Duan, K.; Song J. Analysis for temporal gene expressions under multiple biological conditions. *Stat. Biosci.* **2012**, *4*, 282–299. [CrossRef]
2. Draghici, S.; Kulaeva, O.; Hoff, B.; Petrov, A.; Shams, S.; Tainsky, M.A. Noise sampling method: An ANOVA approach allowing robust selection of differentially regulated genes measured by DNA microarrays. *Bioinformatics* **2003**, *19*, 1348–1359. [CrossRef]
3. Eisen, M.B.; Spellman, P.T.; Brown, P.O.; Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 14863–14868. [CrossRef]
4. Li, H.; Luan, Y.; Hong, F.; Li, Y. Statistical methods for analysis of time course gene expression data. *Front. Biosci.* **2002**, *7*, a90–a98. [CrossRef] [PubMed]
5. Yeung, K.Y.; Ruzzo, W.L. Principal component analysis for clustering gene expression data. *Bioinformatics* **2001**, *17*, 763–774. [CrossRef]
6. Kerr, M.K.; Martin, M.; Churchill, G.A. Analysis of variance for gene expression microarray data. *J. Comput. Biol.* **2000**, *7*, 819–837. [CrossRef] [PubMed]
7. Storey, J.D.; Tibshirani, R. Statistical methods for identifying differentially expressed genes in DNA microarrays. *Methods Mol Biol.* **2003**, *224*, 149–157. [CrossRef]
8. Tusher, V.G.; Tibshirani, R.; Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5116–5121. [CrossRef] [PubMed]
9. Deng, D.; Jahromi, K.R.; Zhou, Z. Influence of biological conditions to temporal gene expression based on variance analysis. In *JSM Proceedings*; American Statistical Association: Alexandria, VA, USA, 2017; pp. 786–800.
10. Deng, D.; Fang, H.-B.; Jahromi, K.R.; Song, J.; Tan, M. Detection of threshold points for gene expressions under multiple biological conditions. *Stat. Interface* **2017**, *10*, 643–655. [CrossRef]
11. Huang, H.; Lee, T.-H. Forecasting Value-at-Risk Using High-Frequency Information. *Econometrics* **2013**, *1*, 127–140. [CrossRef]
12. Gallardo, D.I.; Bourguignon, M.; Galarza, C.E.; Gómez, H.W. A Parametric Quantile Regression Model for Asymmetric Response Variables on the Real Line. *Symmetry* **2020**, *12*, 1938. [CrossRef]
13. Jung, C.; Lee, Y.; Lee, J.; Kim, S. Performance Evaluation of the Multiple Quantile Regression Model for Estimating Spatial Soil Moisture after Filtering Soil Moisture Outliers. *Remote Sens.* **2020**, *12*, 1678. [CrossRef]
14. Chen, L.; Galvao, A.; Song, S. Quantile Regression with Generated Regressors. *Econometrics* **2021**, *9*, 16. [CrossRef]

15. Donoho, D.L.; Johnstone, J.M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **1994**, *81*, 425–455. [CrossRef]
16. Zhang, J.T. Order-dependent Thresholding with Applications to Regression Splines. In *In-Contemporary Multivariate Analysis and Design of Experiments*; World Scentific Publishing Co. Pte. Ltd.: Singapore. 2005; pp. 397–425.
17. Geraci, M.; Bottai, M. Linear quantile mixed models. *Stat. Comput.* **2014**, *24*, 461–479. [CrossRef]
18. Duan, K.; Dammel, C.; Stein, J.; Rabin, H.; Surette, M.G. Modulation of Pseudomonas aeruginosa gene expression by host microflora through interspecies communication. *Mol. Microbiol.* **2003**, *50*, 1477–1491. [CrossRef] [PubMed]