Randomized Quantile Residual for Assessing Generalized Linear Mixed Models with Application to Zero-inflated Microbiome Data

Longhai Li

Department of Mathematics and Statistics
University of Saskatchewan
Saskatoon, SK, CANADA

5 June 2018
Annual Meeting of Statistical Society of Canada
McGill University, Montreal, Canada

Acknowledgements

- This talk is based on the results of the M.Sc thesis project undertaken by **Wei Bai**, co-supervised with **Cindy X. Feng** (U of S).

- Thank Prof. **Wei Xu** (U of T) for providing the microbiome data for this research.

- Thank NSERC and CFI for providing grants for my research.

- Thank the ICSA Canada Chapter, particularly Prof. **Changbao Wu** (U of Waterloo), for organizing and sponsoring this session.

Outline

Section 1

Introduction

Introduction

- The operational taxonomic unit (OTU) counts in microbiome datasets have characteristics of zero-inflation and over-dispersion. Various generalized mixed models have been proposed to to fit the data.
- Correctness in model specification plays extremely important role in statistical inference, for example in calculating p-values/q-values for selecting OTUs that are related to a phenotype.
- Pearson and deviance residuals are often used in practice without justification. However, when applied to count data, the distributions of these residuals are *far* from the normal distribution.
- Randomized quantile residual (RQR) was originally proposed by Dunn and Smyth (1996) as an alternative for Pearson and deviance residuals. However, it has NOT been used much by statisticians.
- We investigate the performance of RQR in checking zero-inflated/modified generalized linear mixed effect (GLMM) models using simulated and real datasets.

Section 2

Zero-inflated/modified Generalized Linear Mixed Models

Generalized Linear Mixed Model

As an example of GLMM, NB mixed model (NB) is described as follows:

- A probability distribution for the response (y_i) given a mean function μ_i and other parameters, eg.

$$y_i|\mu_i \sim \text{Negative-Binomial}(\mu_i, k)$$

- A link function for linking the mean μ_i to a linear function of fixed factor (X_i) and random factors (Z_i), eg.

$$\log(\mu_i) = X_i\beta + Z_i u$$

- Certain penalization (often normal) is imposed to u.

Zero-inflated Poisson (ZIP) Model: I

- The zero-inflated Poisson with parameters λ_i and p_i, denoted by $ZIP(\lambda_i, p_i)$, is defined as:

$$y_i \sim \begin{cases} 0, & \text{with probability } p_i \\ \text{Poisson}(\mu_i), & \text{with probability } 1 - p_i. \end{cases} \tag{1}$$

- The following link functions are often used:

$$\log(\mu_i) = \text{offset}_i + X_i\beta + Z_i u \tag{2}$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \widetilde{\text{offset}}_i + \tilde{X}_i\tilde{\beta} + \tilde{Z}_i\tilde{u}, \tag{3}$$

Zero-inflated Poisson (ZIP) Model: II

- The PMF and CDF of the ZIP distribution:

$$\text{dzip}(y_i = 0) = p_i + (1 - p_i) \times e^{-\mu_i} \tag{4}$$

$$\text{dzip}(y_i = j) = (1 - p_i)\frac{e^{-\mu_i}\mu_i^j}{j!}, \text{ for } j > 0 \tag{5}$$

$$\text{pzip}(y_i = J; \mu_i, p_i) = p_i + (1 - p_i)\text{ppois}(J, \mu_i). \tag{6}$$

- The mean and variance of a ZIP random variable can be calculated by

$$E(y_i) = (1 - p_i) \times \mu_i \tag{7}$$

$$V(y_i) = (1 - p_i) \times \left(\mu_i + p_i \times {\mu_i}^2\right). \tag{8}$$

Zero-inflated Negative-Binomial (ZINB) Model: I

Zero-inflated NB(ZINB) can be defined similarly as ZIP, with Poisson replaced by NB.

- The PMF and CDF of the ZINB distribution:

$$\text{dzinb}(y_i = 0) = p_i + (1 - p_i) \times \left(\frac{k}{k + \mu_i}\right)^k \tag{9}$$

$$\text{dzinb}(y_i = j) = (1 - p_i) \times \text{dnb}(j, \mu_i, k), \text{ for } j > 0 \tag{10}$$

$$\text{pzinb}(y_i; \mu_i, k, p_i) = p_i + (1 - p_i)\text{pnb}(y_i, \mu_i, k) \tag{11}$$

- The mean and variance of a ZIP random variable can be calculated by

$$E(y_i) = (1 - p_i) \times \mu_i \tag{12}$$

$$V(y_i) = (1 - p_i) \times \left(\mu_i + \frac{{\mu_i}^2}{k}\right) + {\mu_i}^2 \times ({p_i}^2 + p_i) \tag{13}$$

Zero-modified Poisson (ZMP): I

- Zero-Modified Model: Zero-modified models are also called hurdle models. A logistic regression for the zero indicator (Z_i):

$$Pr(Z_i = z) = \begin{cases} \pi_i, & z = 0 \\ 1 - \pi_i, & z = 1. \end{cases} \tag{14}$$

- Given Z_i, the conditional probability mass function for Y_i is:

$$\begin{cases} Pr(Y_i = y_i | Z_i = 0) = I(y_i = 0) \\ Pr(Y_i = y_i | Z_i = 1) = \frac{\text{dpois}(y_i)}{1 - \text{dpois}(0)} I(y_i > 0). \end{cases} \tag{15}$$

- The unconditional probability mass function for Y_i is

$$Pr(Y_i = y_i) = \begin{cases} \pi_i, & \text{if } y_i = 0 \\ (1 - \pi_i) \frac{\text{dpois}(y_i)}{1 - \text{dpois}(0)}, & \text{if } y_i > 0. \end{cases} \tag{16}$$

Zero-modified Poisson (ZMP): II

- We often used the log link functions for non-zero count mean μ_i and logistic link for π_i:

$$\log(\mu_i) = \text{offset}_i + X_i\beta + Z_i u \tag{17}$$

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \widetilde{\text{offset}}_i + \tilde{X}_i\tilde{\beta} + \tilde{Z}_i\tilde{u}, \tag{18}$$

- The PMF and CDF of ZMP distribution:

$$\text{dzmp}(y_i = 0) = \pi_i \tag{19}$$

$$\text{dzmp}(y_i = j) = (1-\pi_i)\frac{\text{dpois}(j)}{1-\text{ppois}(0)}, \text{ for } j > 0 \tag{20}$$

$$\text{pzmp}(y_i; \mu_i, \pi_i) = \pi_i + (1-\pi_i)\frac{\text{ppois}(y_i; \mu_i, \pi_i) - \text{ppois}(0)}{1-\text{ppois}(0)}, \tag{21}$$

Zero-modified NB (ZMNB)

- ZMNB model can be defined analogously. The PMF and CDF of ZMNB distribution:

$$\text{dzmnb}(y_i = j) = \pi_i I(j = 0) + (1 - \pi_i)\frac{\text{dnb}(y_i)}{1 - \text{pnb}(0)} I(j > 0) \tag{22}$$

$$\text{pzmnb}(y_i; \mu_i, k, \pi_i) = \pi_i + (1 - \pi_i)\frac{\text{pnb}(y_i) - \text{pnb}(0)}{1 - \text{pnb}(0)}. \tag{23}$$

- The same link functions as in ZMP are used for ZMNB.
- The mean and variance of ZMNB:

$$E(y_i) = \frac{1 - \pi_i}{1 - p_0} \times \mu_i \tag{24}$$

$$V(y_i) = \frac{1 - \pi_i}{1 - p_0} \times \left(\mu_i + \mu_i^2 + \frac{{\mu_i}^2}{k}\right) - \left(\frac{1 - \pi_i}{1 - p_0} \times \mu_i\right)^2. \tag{25}$$

Section 3

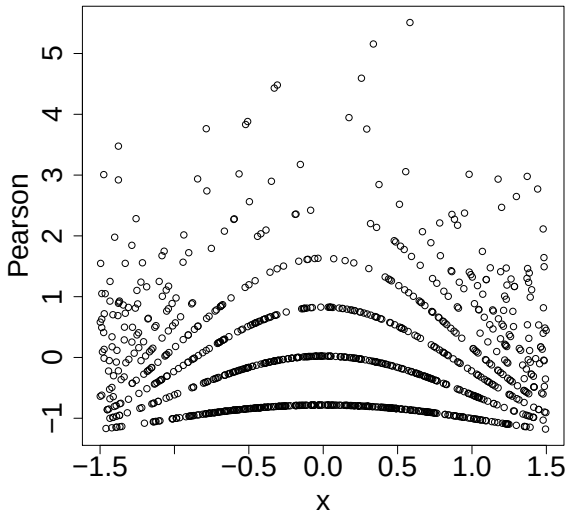
Randomized Quantile Residual

Problems with Pearson and Deviance Residuals

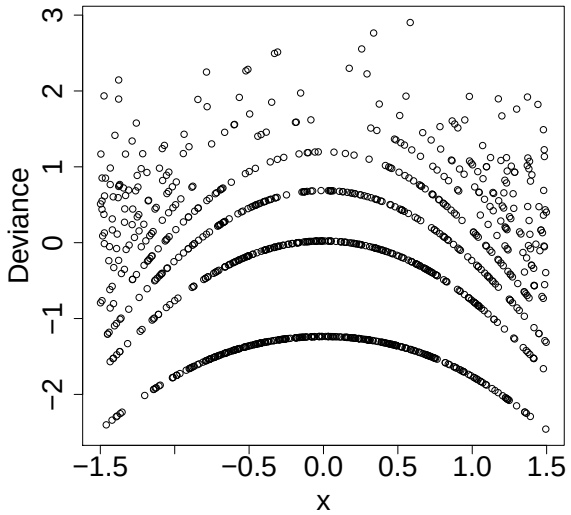
- In regression models for discrete outcomes, the residuals are far from normality, with residuals clustering on lines according to distinct response values, which poses great challenges for visual inspection. Therefore, residual plots for the diagnosis of models for discrete outcome variables give very limited meaningful information for model diagnosis.
- The *Pearson* χ^2 *statistic* is written as, $X^2 = \sum_{i=1}^{n} r_i^2$, and the *deviance* ($\chi^2$ *statistic*) is written as, $D = \sum_{i=1}^{n} d_i^2$. The asymptotic distribution of D and X^2 under the true model is often assumed to be χ^2_{n-p}. However, the use of this asymptotic distribution for both X^2 and D is lack of theoretical underpinning.

A First Look at Three Residuals

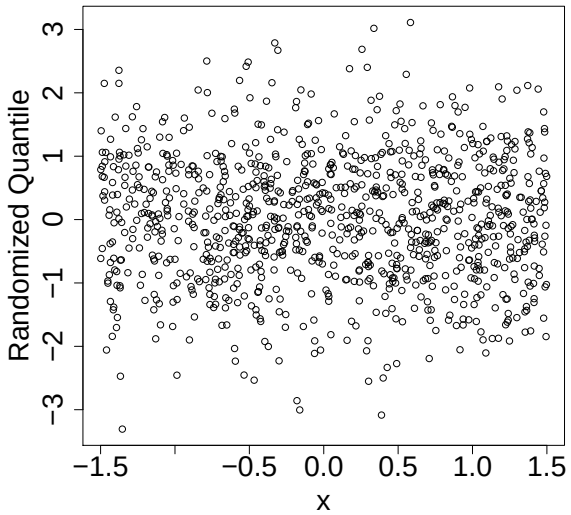
Pearson



Deviance



Randomized Quantile



- A simulated dataset is checked against the **true** generating model. However, Pearson and deviance residuals exhibit trend and cluster in lines.
- In addition, the often used χ^2 tests are not well-calibrated.

Definition of Randomized Quantile Residual

- Predictive p-value for continuous y_i:

$$F(y_i; \hat{\mu}_i, \hat{\phi}) = P(Y_i \leq y_i \mid \hat{\mu}_i, \hat{\phi})$$

- Randomized predictive p-value
  If F is discrete, the estimated lower tail probability is randomized into a uniform random number.

$$F^*(y_i; \hat{\mu}_i, \hat{\phi}, u_i) = F(y_i-; \hat{\mu}_i, \hat{\phi}) + u_i\, P(y_i; \hat{\mu}_i, \hat{\phi}), \tag{26}$$

  where u_i from uniform distribution on $(0, 1]$, $F(y_i-; \hat{\mu}_i, \hat{\phi})$ is the lower limit of F at y_i, i.e., $\sup_{y<y_i} F(y; \hat{\mu}_i, \hat{\phi})$, the lower limit in the "gap" of $F(\cdot, \hat{\mu}_i, \hat{\phi})$ at y_i.
- Randomized quantile residual

$$q_i = \Phi^{-1}(F^*(y_i; \hat{\mu}_i, \hat{\phi}, u_i)) \tag{27}$$

  where Φ^{-1} is the quantile function of a standard normal distribution

An Illustrative Example for RQR: I

- **The true model:**
  We simulate a response variable of size $n = 1000$ from a Poisson model with

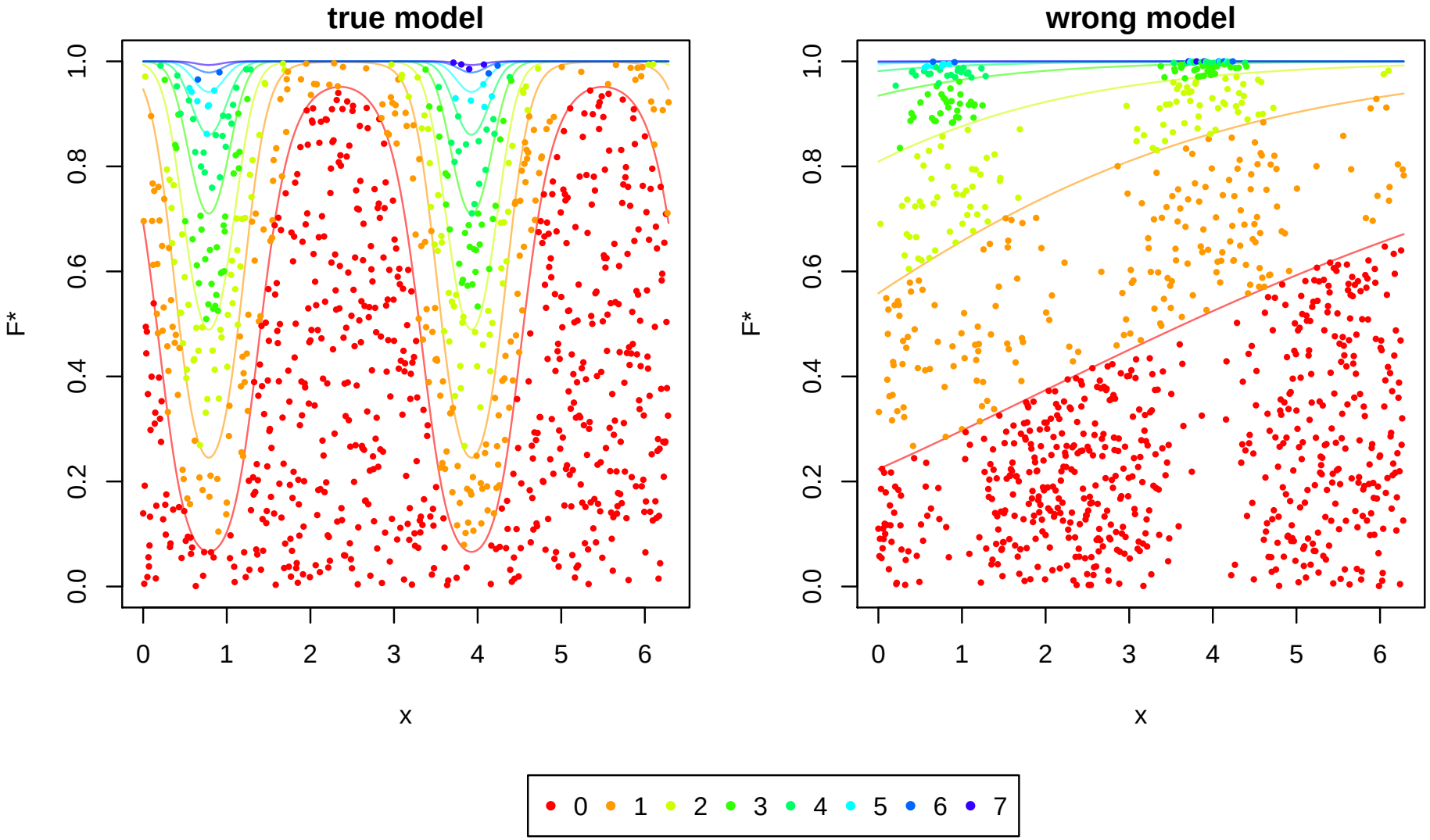
$$\log(\mu_i) = -1 + 2\sin(2x_i),$$

  where μ_i is the expected mean count for the ith subject and $x_i \sim Uniform(0, 2\pi)$, $i = 1, \cdots, n$
- **A wrong model:**
  Poisson model with mean structure

$$\log(\mu_i) = \beta_0 + \beta_1 x_i$$

  with x_i as a predictor with linear effect.

An Illustrative Example for RQR: II



Normality of Randomized Quantile Residual (RQR)

Theorem

Suppose a continuous random variable Y has the CDF $F(y)$, then $F(Y)$ is uniformly distributed on (0,1].

Theorem

Suppose the true distribution of Y_i given X_i has the CDF $F(y_i; \mu_i, \phi)$ and PMF $P(y_I; \mu_i, \phi)$, where μ_i is a function of X_i involving the model parameters. The randomized lower tail probability $F^*(y_i; \mu_i, \phi, u_i)$ is defined as $F(y_i-; \mu_i, \phi) + u_i\, P(y_I; \mu_i, \phi)$ (26). Suppose U_i is uniformly distributed on (0,1]. Then, we have

$$F^*(Y_i; \mu_i, \phi, U_i) \sim \mathit{Uniform}((0, 1]), \tag{28}$$

and

$$q_i = \phi^{-1}(F^*(Y_i; \mu_i, \phi, U_i)) \sim N(0, 1). \tag{29}$$

Proof of Normality of RQR

For any interval $B \subseteq (0, 1]$,

$$P(F^*(Y_i; \mu_i, \phi, U_i) \in B | Y_i = k^{(j)}) = \frac{\text{length}(F^{(j)} \cap B)}{p^{(j)}},$$

where length$(\cdot)$ is the length of interval. By the law of total probability,

$$\begin{aligned}
& P(F^*(Y_i; \mu_i, \phi, U_i) \in B) && (30)\\
= & \sum_{j=1}^{\infty} P(F^*(Y_i; \mu_i, \phi, U_i) \in B | Y_i = k^{(j)}) \times P(Y_i = k^{(j)}) && (31)\\
= & \sum_{j=1}^{\infty} \frac{\text{length}(F^{(j)} \cap B)}{p^{(j)}} \times p^{(j)} && (32)\\
= & \sum_{j=1}^{\infty} \text{length}(F^{(j)} \cap B) && (33)\\
= & \ \text{length}(\cup_{j=1}^{\infty} F^{(j)} \cap B) = \text{length}(B) && (34)
\end{aligned}$$

Section 4

Simulation Studies

Subsection 1

Description of Data Generating Process

General Form of Microbiome Dataset

	OTU_1 ... OTU_m	Total Reads	Host Factors	Sample Variables
	Y_1 Y_m	Offset	Fixed Factors	Random Factors
sample 1	Y_{11} ... Y_{1m}	T_1	X_{11} ... X_{1s}	Z_{11} ... Z_{1t}
.	. ... .	.	...	...
.	. ... .	.	...	...
.	. ... .	.	...	...
sample n	Y_{n1} ... Y_{nm}	T_n	X_{n1} ... X_{ns}	Z_{n1} ... Z_{nt}

Link Functions and Parameters in Data Generation

- Link Functions

$$\log(\mu_i) = \log(T_i) + \beta_0 + \beta^{(1)}_{X_{i1}} + ... + \beta^{(s)}_{X_{is}} + u^{(1)}_{Z_{i1}} + ... + u^{(t)}_{Z_{it}},$$
$$\log(\frac{\pi_i}{1-\pi_i}) = \tilde{\beta}_0 + \tilde{\beta}^{(1)}_{X_{i1}} + ... + \tilde{\beta}^{(s)}_{X_{is}} + \tilde{u}^{(1)}_{Z_{i1}} + ... + \tilde{u}^{(t)}_{Z_{it}}$$

- Parameters:

Parameter	Generator
$\tilde{\beta}_0$	-0.2
$\beta, \tilde{\beta}$	$N(0, 0.1^2)$
$u, \tilde{u}$	$N(0, 2^2)$
k (ZMNB)	Unif(1,2)
T_i	Poisson(3×10^5)

- Other Settings: $m = 3000, s = 3, t = 3$; each fixed factor has 5 levels and each random factor has 10 levels.

Steps to Generate OTUs with ZMP/ZMNB Model

Step 1: Generate matrix of fixed and random factors, and total reads T_i randomly (used for all Y_j).
For each response Y_j:
Step 2: Compute π_{ij} and μ_{ij} using link functions with randomly generated parameters.
Step 3: We generate a count indicator Z_{ij} as a binary Bernoulli random variable:

$$Z_{ij} = \begin{cases} 0, & \text{with probability } \pi_{ij} \\ 1 & \text{with probability } 1 - \pi_{ij}. \end{cases} \tag{35}$$

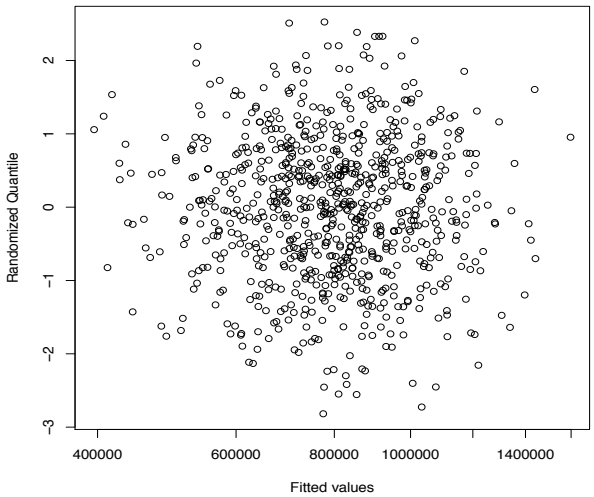
Step 4: If the indicator $Z_{ij} = 0$, then $Y_{ij} = 0$. If the indicator $Z_{ij} = 1$, then Y_{ij} follows a truncated Poisson or NB model, e.g.,

$$Y_{ij} \sim \text{Truncated-Poisson}(\mu_{ij}).$$

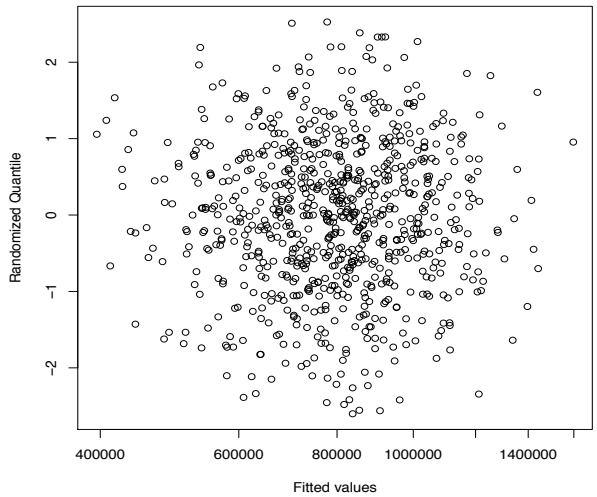
Subsection 2

Assessing Models for Datasets Simulated from ZMP Model

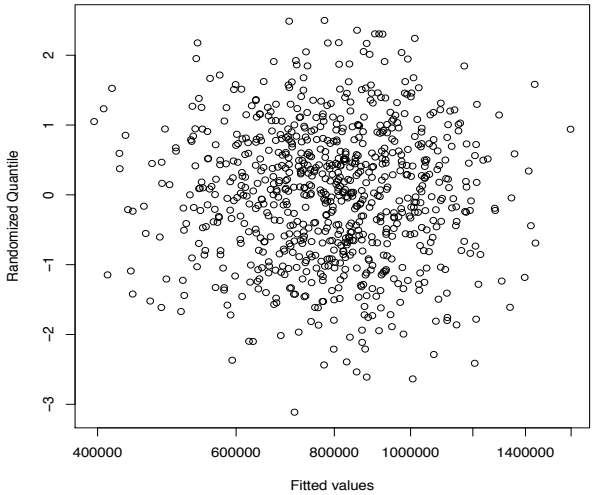
Checking One Y_j: RQR plot vs Fitted Values



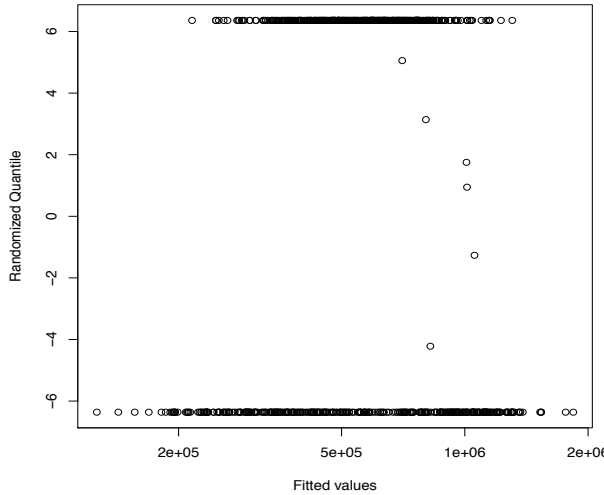
ZMP



ZIP

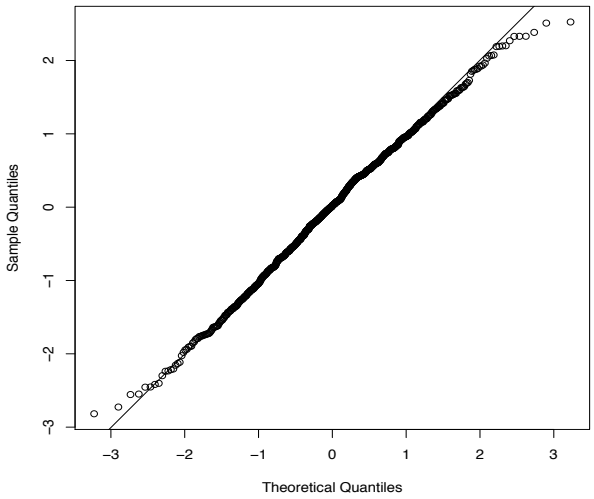


ZMNB

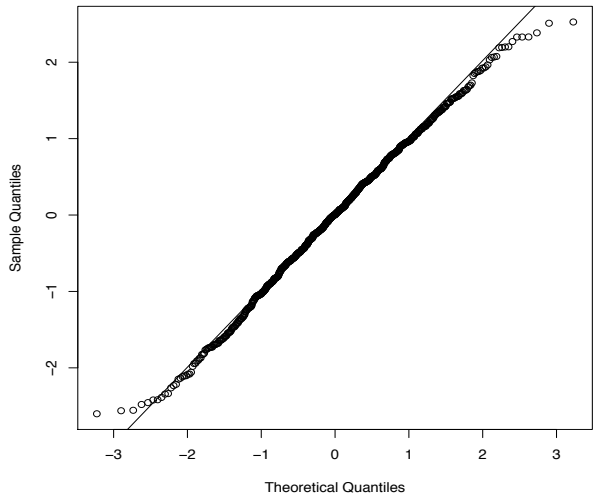


Poisson

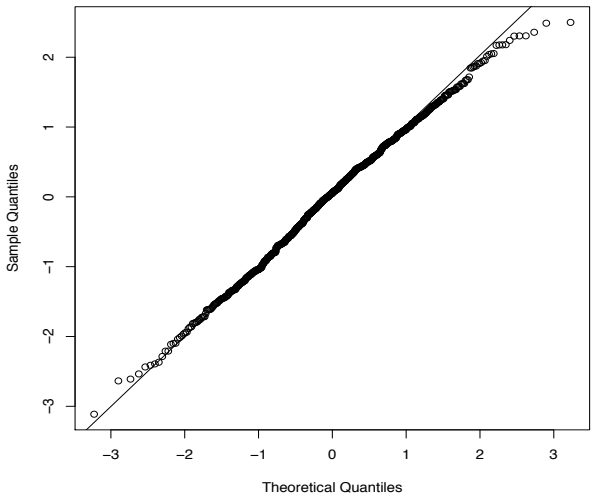
Checking One Y_j: QQ-plot of RQR



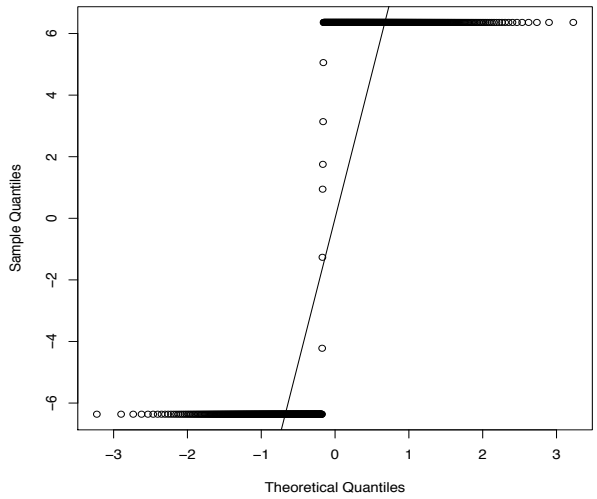
ZMP



ZIP

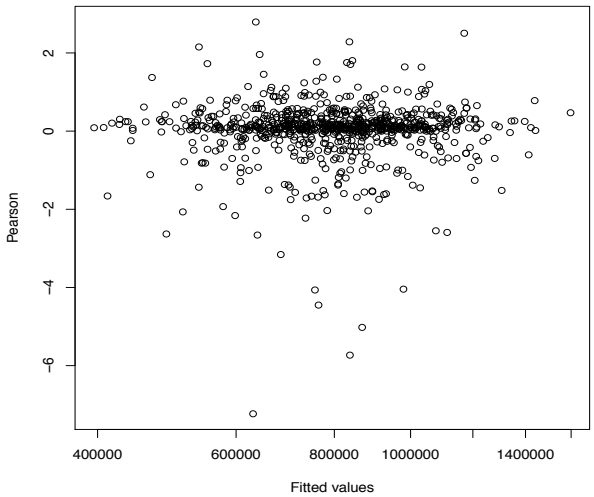


ZMNB

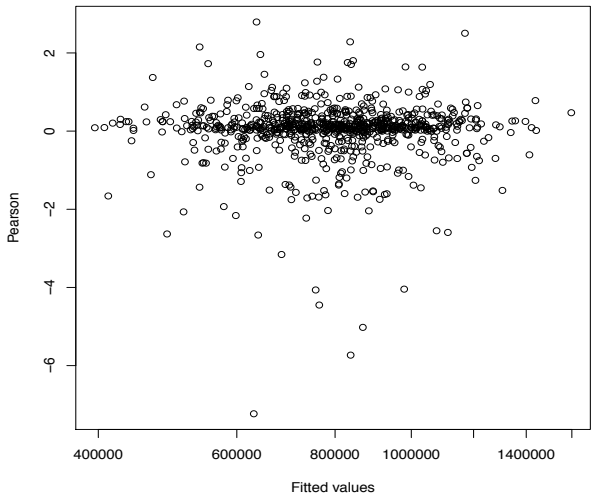


Poisson

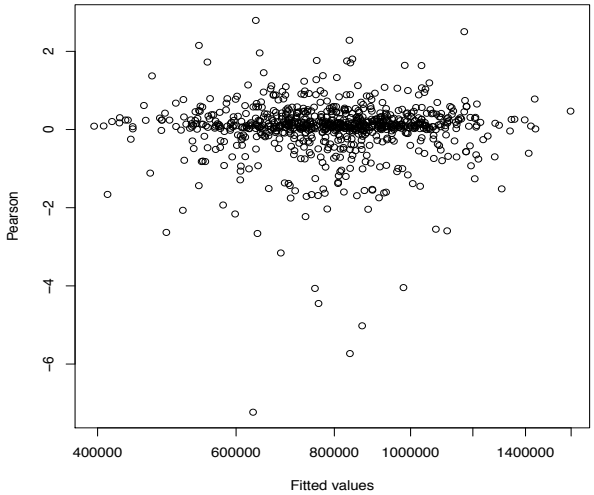
Checking One Y_j: Pearson Residual vs Fitted Values



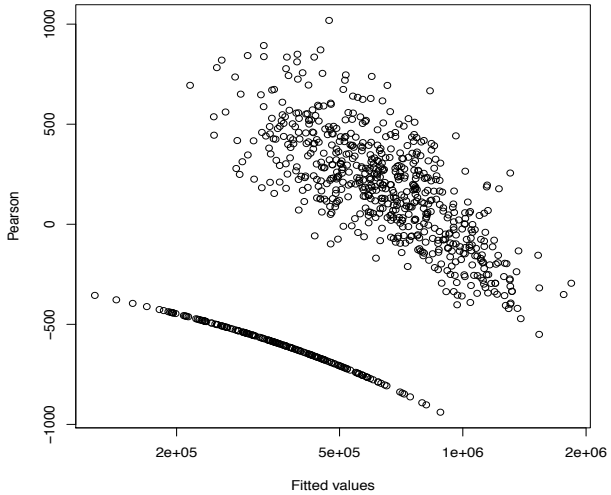
ZMP



ZIP

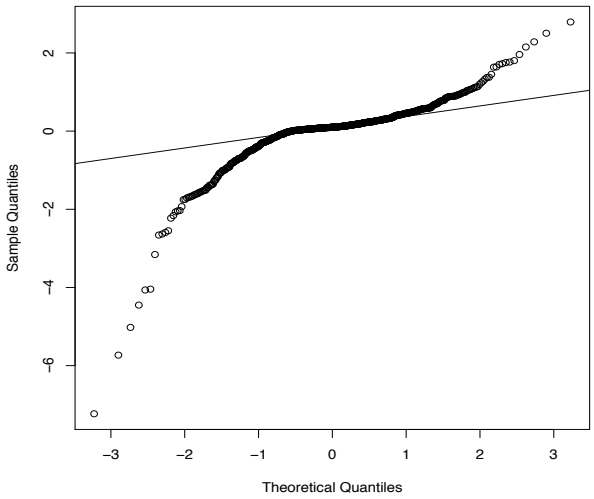


ZMNB

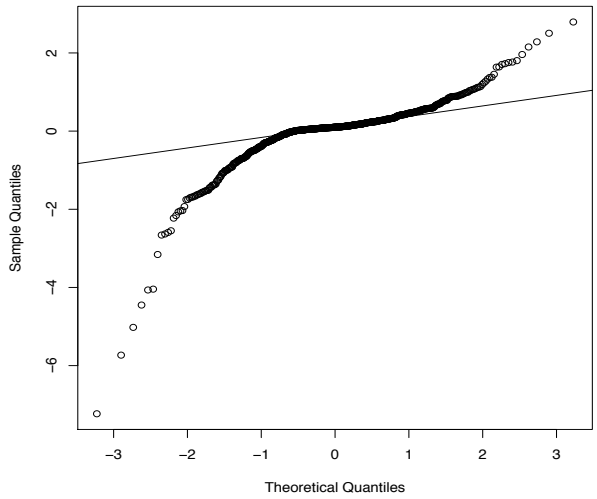


Poisson

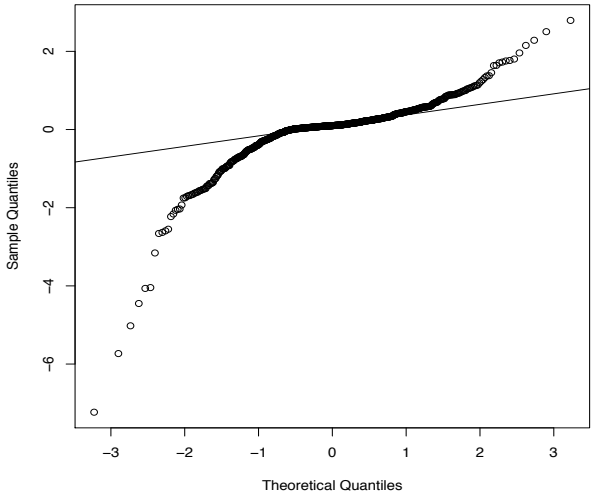
Checking One Y_j: QQ-plot of Pearson Res.



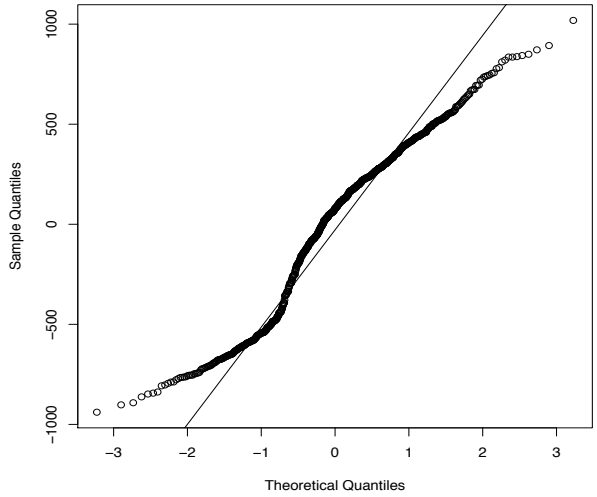
ZMP



ZIP

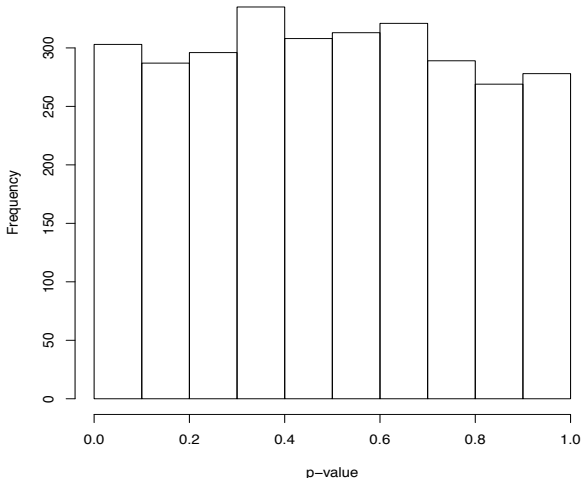


ZMNB

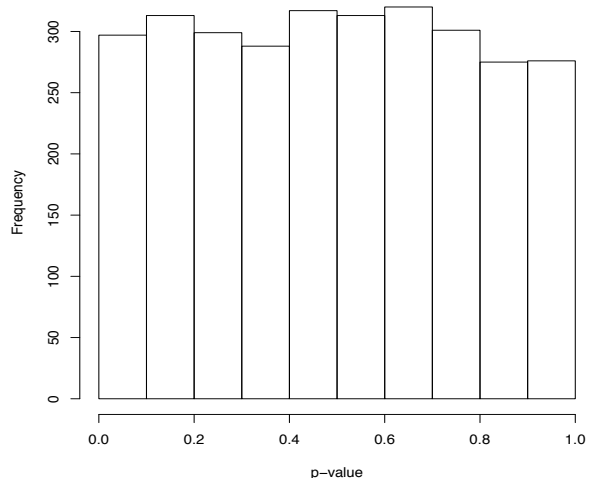


Poisson

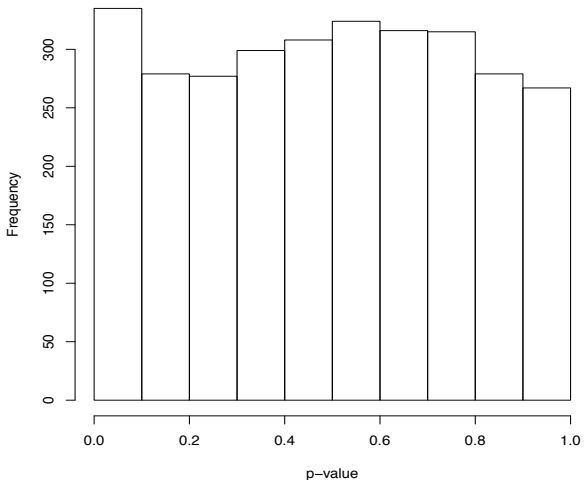
Checking All Y_j's: 3000 Shapiro-Wilk P-values of RQR



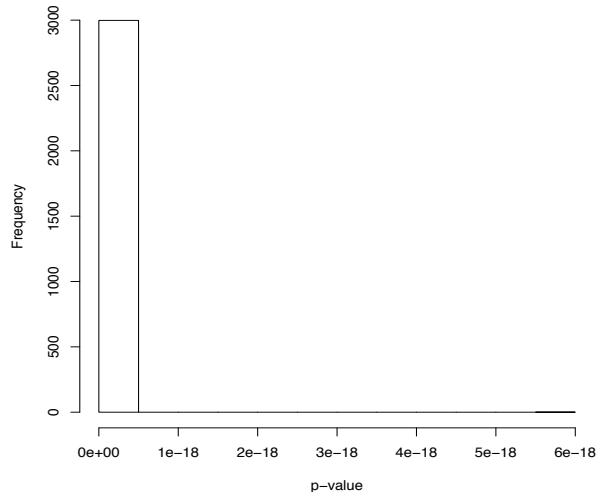
ZMP



ZIP

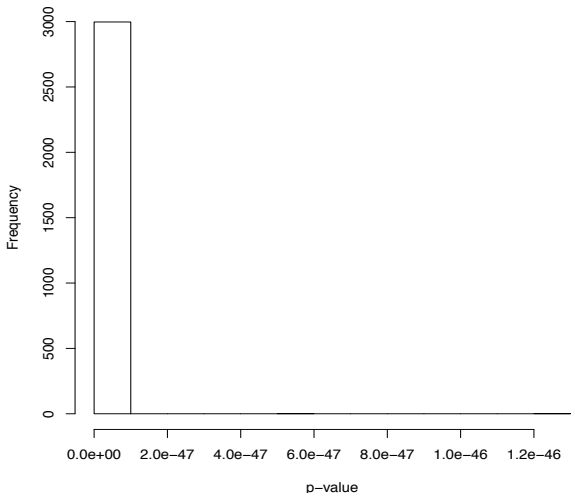


ZMNB

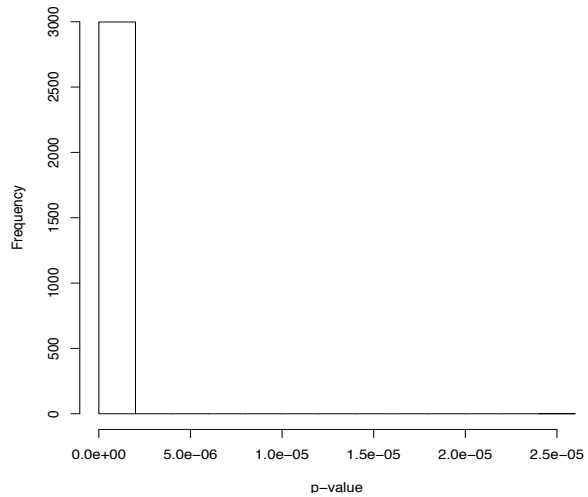


Poisson

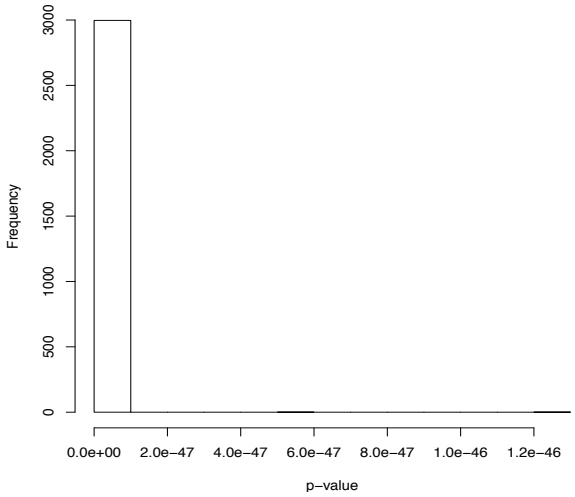
Checking All Y_j's: 3000 Shapiro-Wilk P-values of Pearson



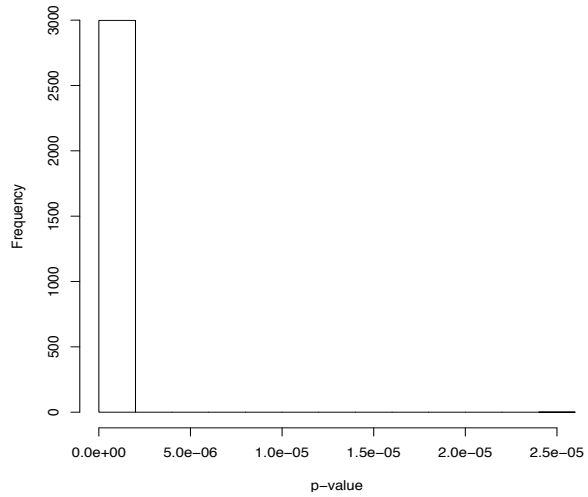
ZMP



ZIP



ZMNB



Poisson

Type 1 Error Rates and Power

Using 0.05 as cutoff, probabilities of rejecting fitted models with RQRs and Pearson residuals in 3000 Y_j's are shown as follows:

Table 1: Using Randomized Quantile Residual

Sample size	ZMP	ZIP	ZMNB	Poisson
200	0.142	0.139	0.145	1.000
400	0.074	0.090	0.102	0.999
800	0.068	0.068	0.082	1.000
1600	0.060	0.061	0.059	1.000
3200	0.051	0.051	0.063	1.000

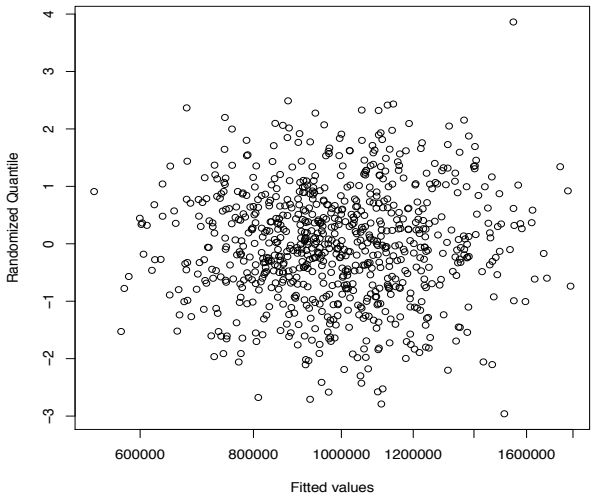
Table 2: Using Pearson Residual

Sample size	ZMP	ZIP	ZMNB	Poisson
200	0.984	0.986	0.983	0.997
400	1.000	1.000	1.000	1.000
800	1.000	1.000	1.000	1.000
1600	1.000	1.000	1.000	1.000
3200	1.000	1.000	1.000	1.000

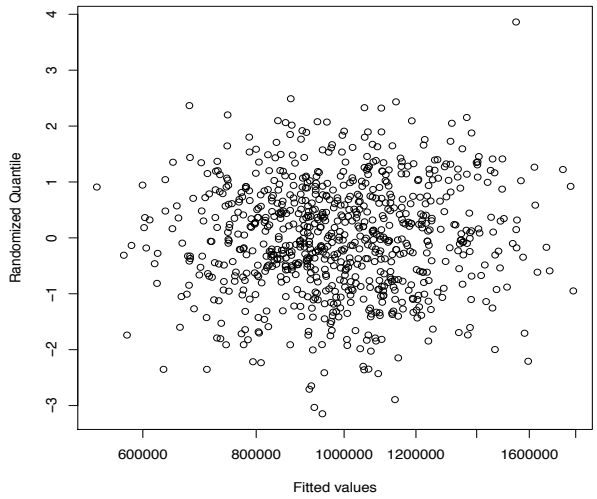
Subsection 3

Assessing Models for Datasets Simulated from ZMNB Model

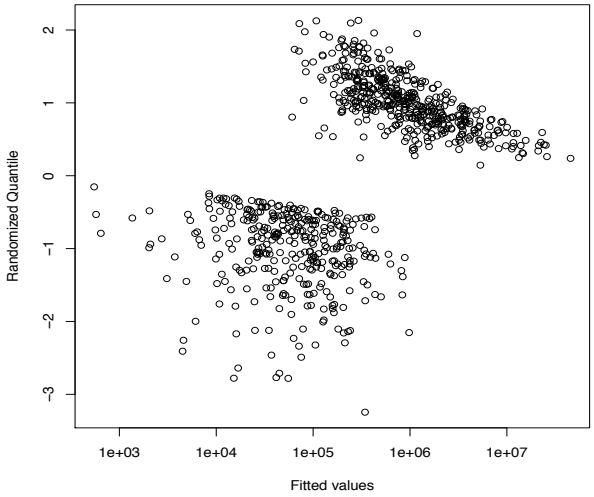
Checking One Y_j: RQR vs Fitted Values



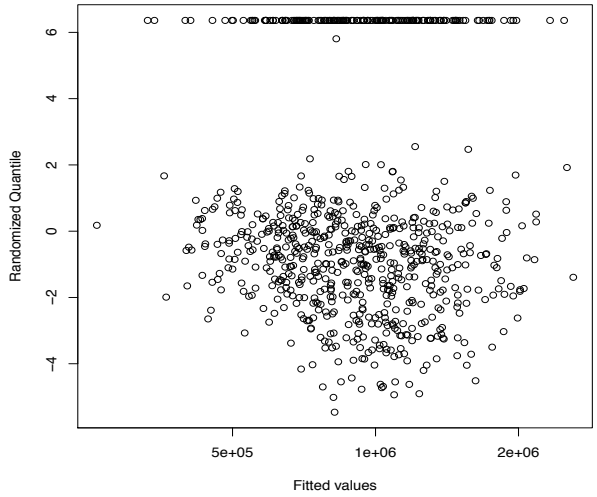
ZMNB



ZINB

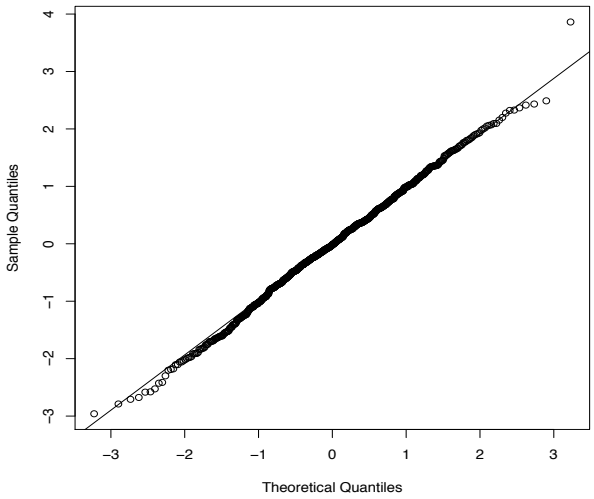


NB

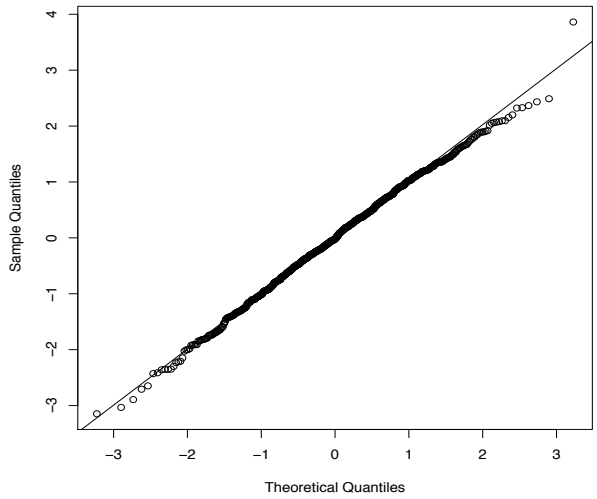


ZMP

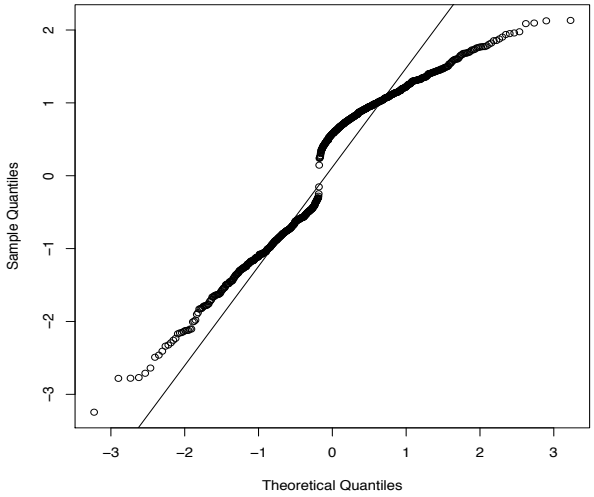
Checking One Y_j: QQ plot of RQR



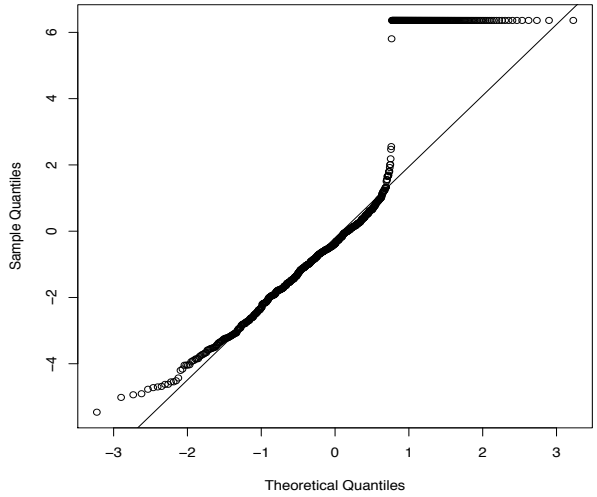
ZMNB



ZINB

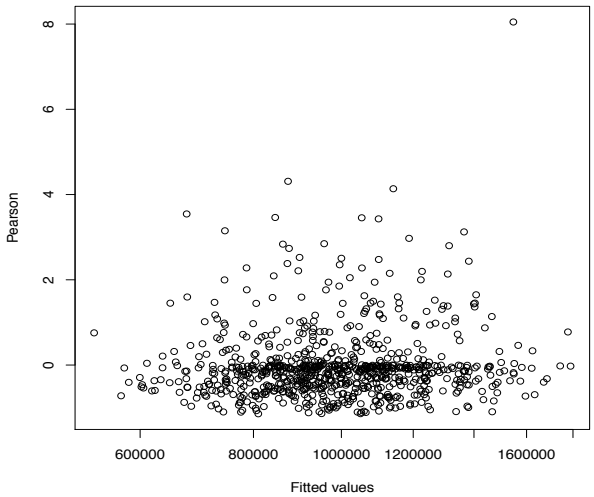


NB

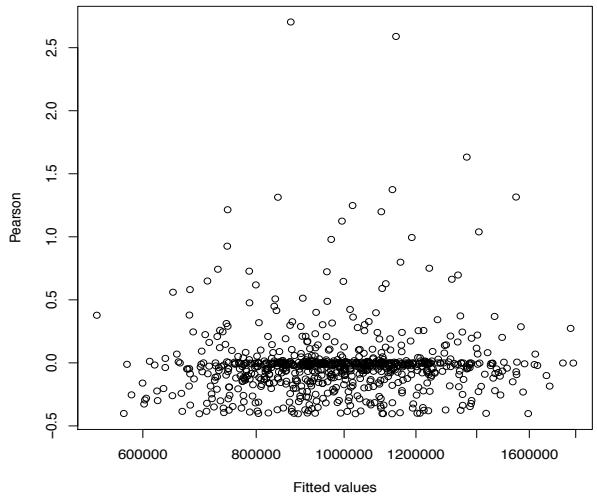


ZMP

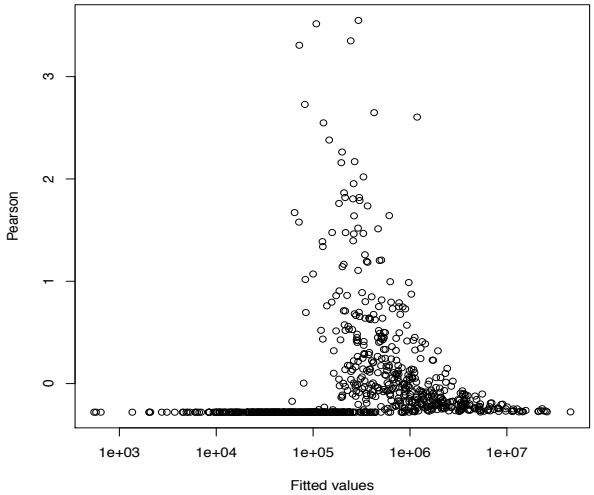
Checking One Y_j: Pearson Residuals vs Fitted Values



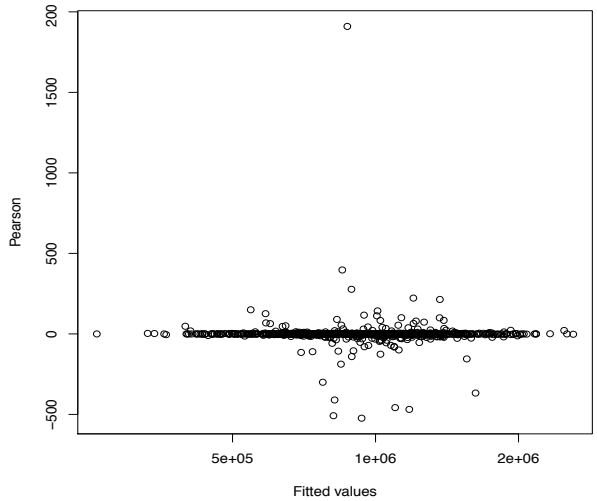
ZMNB



ZINB

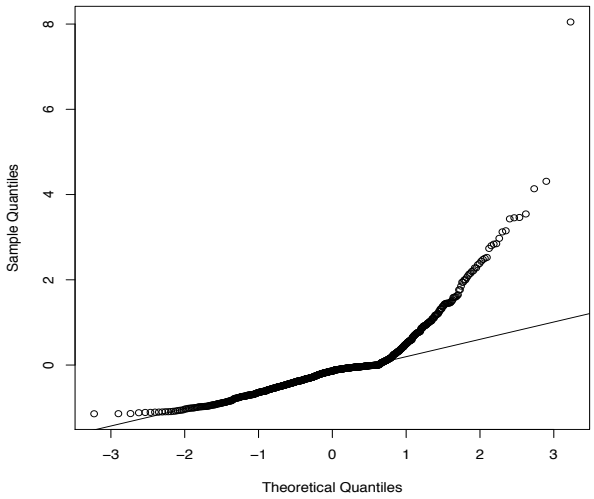


NB

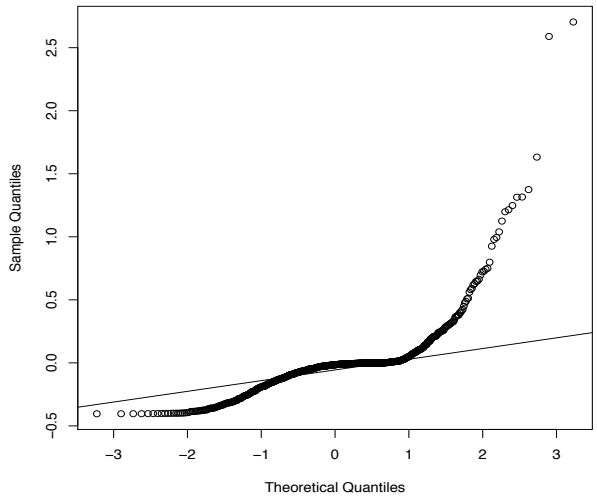


ZMP

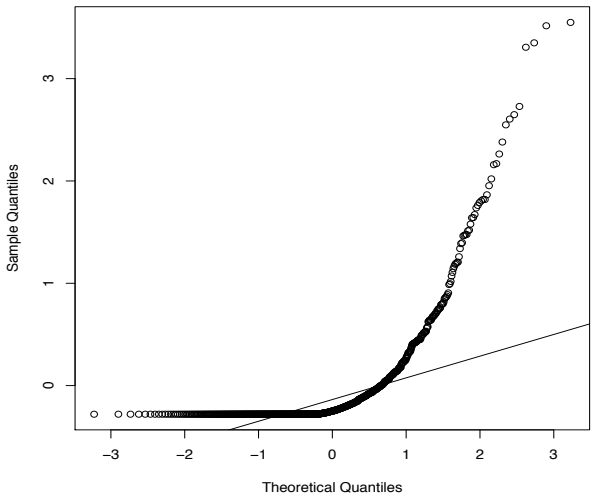
Checking One Y_j: QQ plot of Pearson Residuals



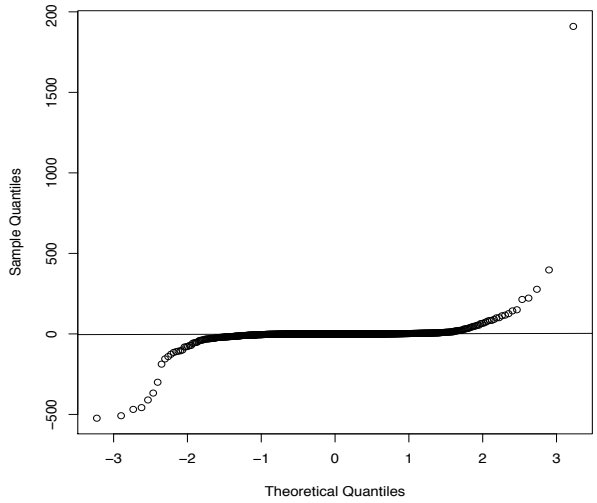
ZMNB



ZINB

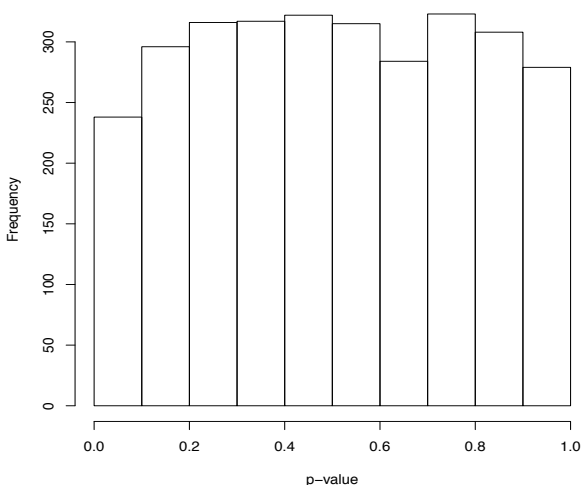


NB

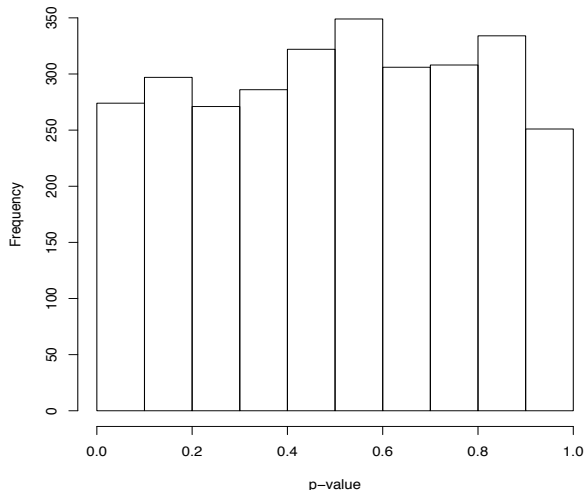


ZMP

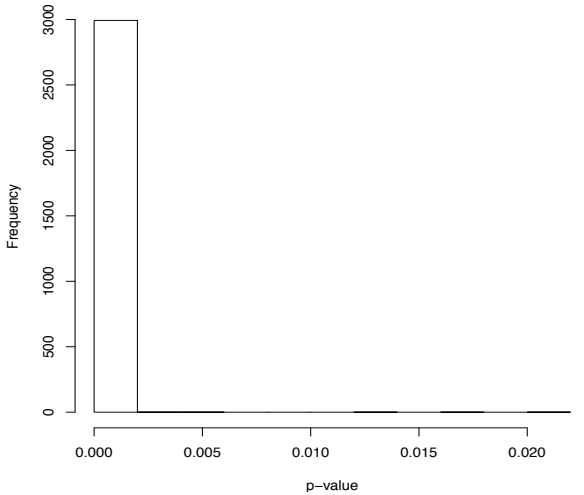
Checking All Y_j's: 3000 Shapiro-Wilk P-values of of RQR



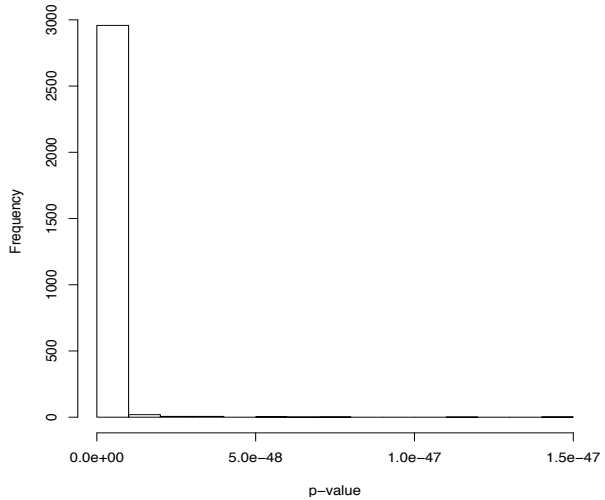
ZMNB



ZINB

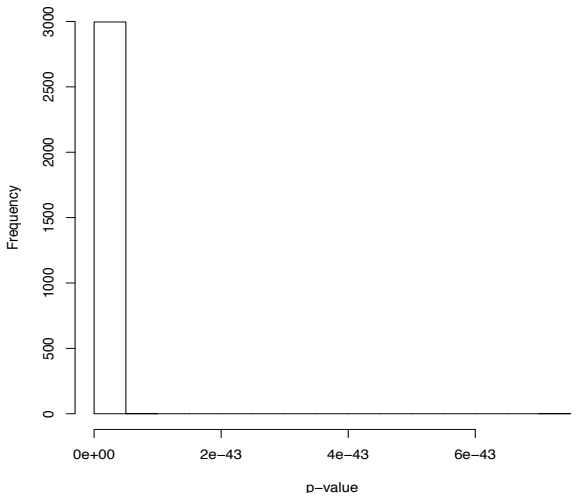


NB

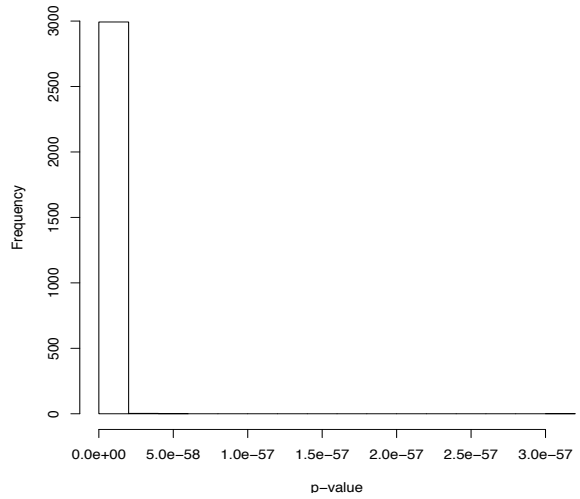


ZMP

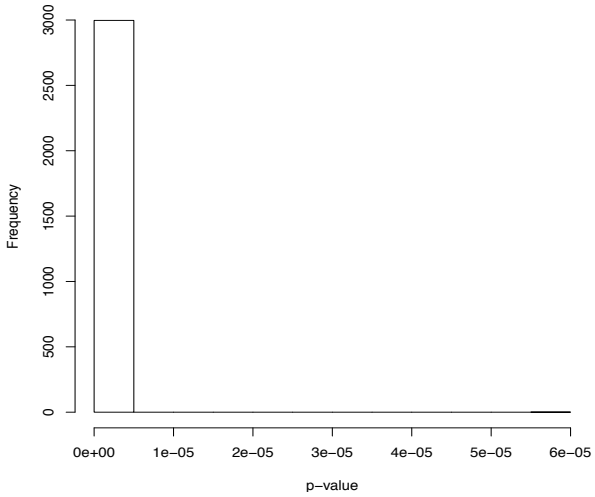
Checking All Y_j's: 3000 Shapiro-Wilk P-values of Pearson



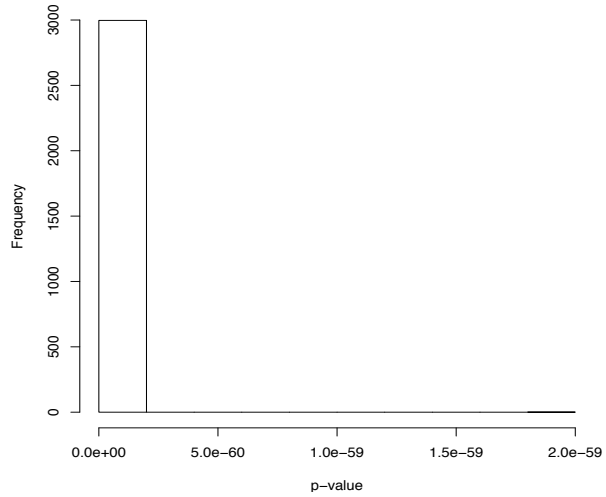
ZMNB



ZINB



NB



ZMP

Type 1 Error Rates and Power

Using 0.05 as cutoff, probabilities of rejecting fitted models with RQRs and Pearson residuals in 3000 Y_j's are shown as follows:

Table 3: Using Randomized Quantile Residuals

Sample size	ZMNB	ZINB	NB	ZMP
200	0.067	0.153	0.957	1.000
400	0.057	0.063	0.883	1.000
800	0.053	0.049	0.759	1.000
1600	0.047	0.055	0.928	1.000
3200	0.040	0.042	1.000	1.000

Table 4: Using Pearson Residuals

Sample size	ZMNB	ZINB	NB	ZMP
200	1.000	1.000	1.000	1.000
400	1.000	1.000	1.000	1.000
800	1.000	1.000	1.000	1.000
1600	1.000	1.000	1.000	1.000
3200	1.000	1.000	1.000	1.000

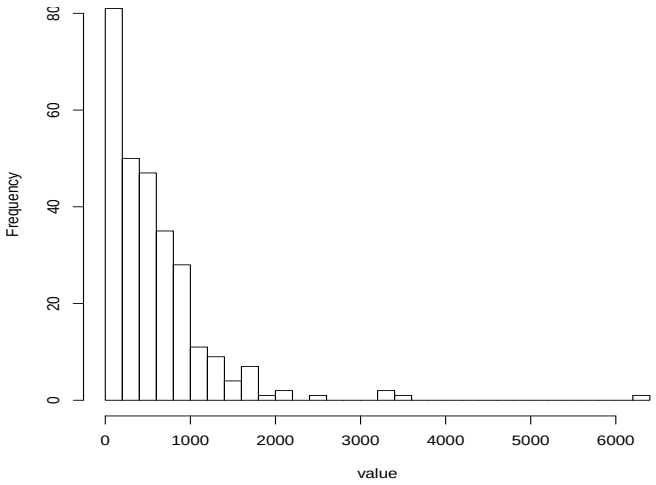
Section 5

Application to a Twin Study OTU Dataset

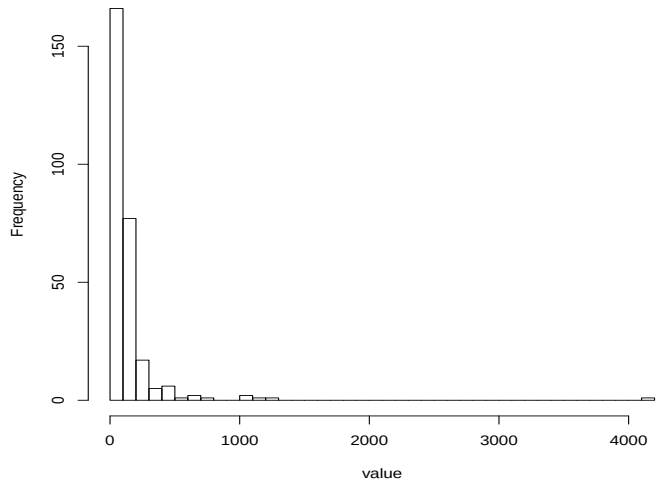
Data Description

- We use a twin study OTU data at the genus level. There are $m = 14$ different genera (14 Y_j) on $n = 287$ samples in total.
- We apply six different models proposed before to fit into this twin study OTU data and use randomized quantile residuals to test the goodness of fit for all OTUs.
- We choose **ancestry** and **obesity** to be host factors while **age** and **family** to be random factors.
- At the genus level, the dataset does not have many zero. However, the ordinary NB and Possion models do not fit the dataset well (to be shown).
- We combine small OTU counts smaller than 10 into a bin called "zero" for 10 genera, and using larger thresholds (less than 150) for other 4 genera.

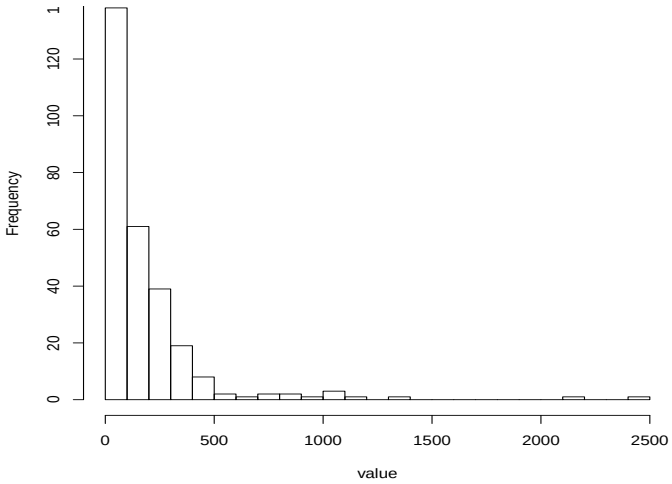
Histograms of OTUs of 4 Genera



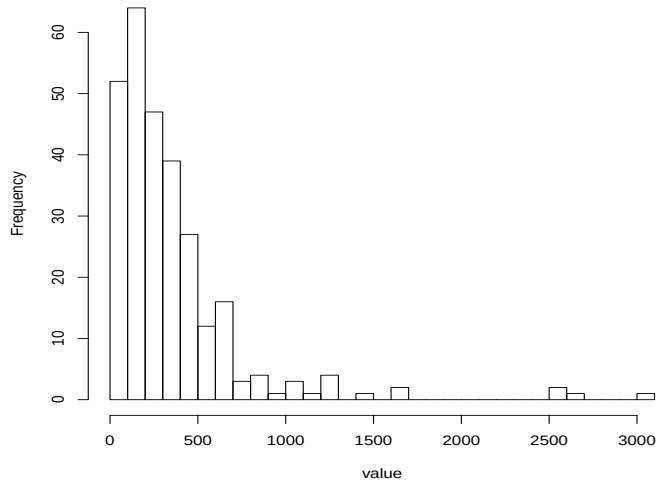
Bacteroides



Lachnospiraceae..g

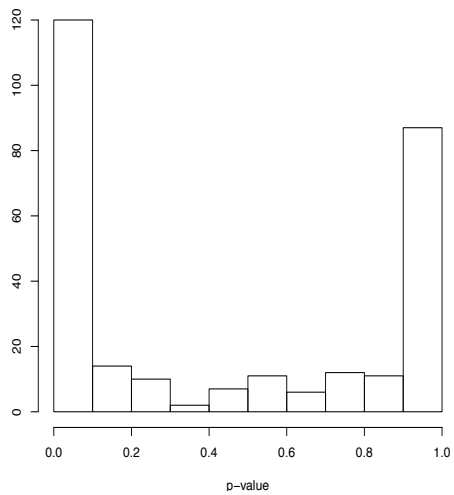


Roseburia

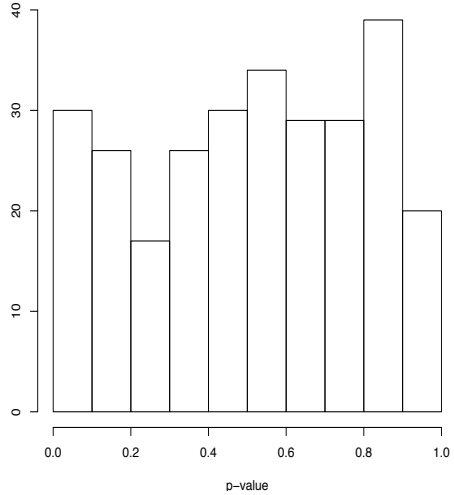


Faecalibacterium

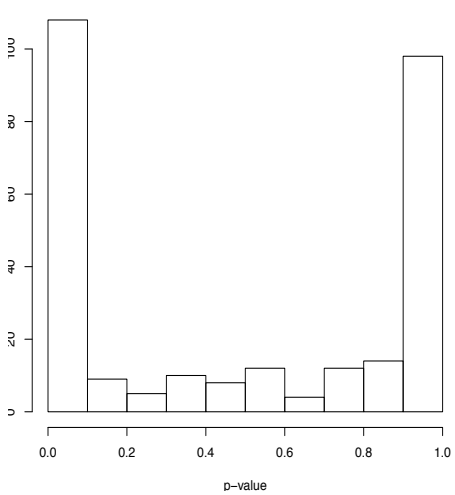
Histograms of Randomized Predictive P-values for "Euba"



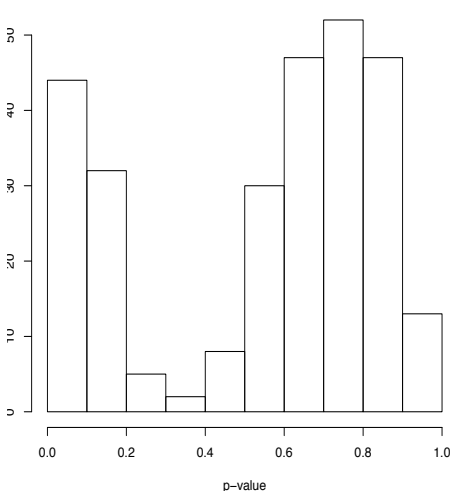
Poisson1



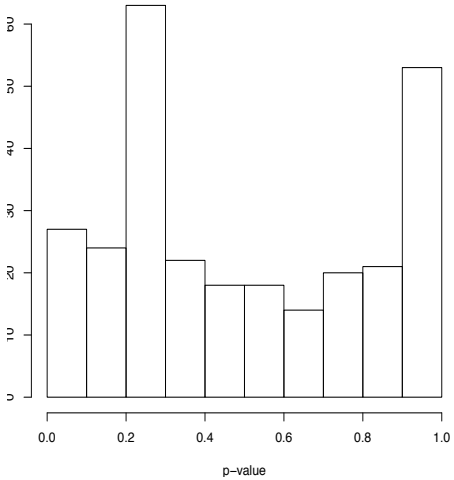
NB1



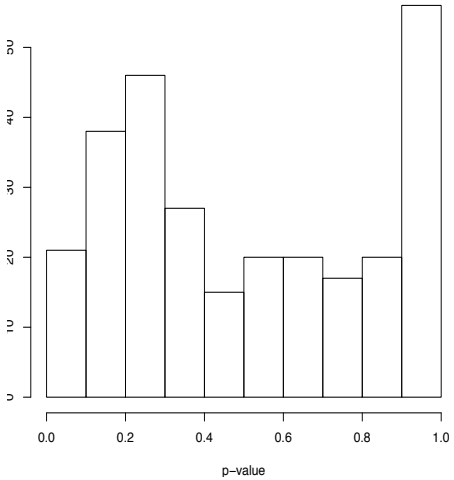
Poisson



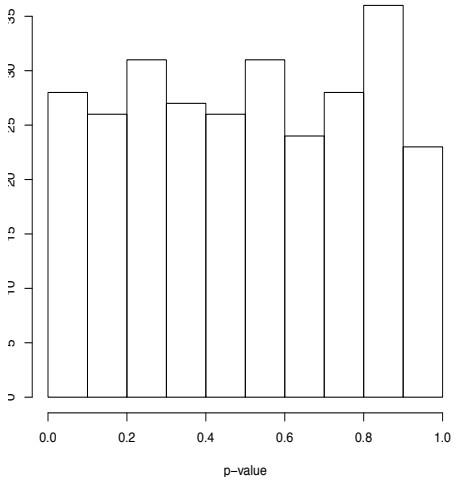
NB



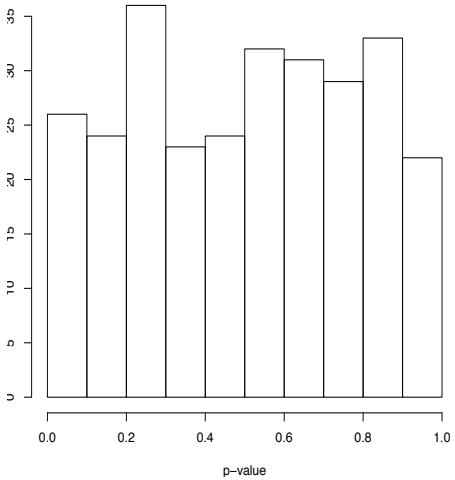
ZIP



ZMP

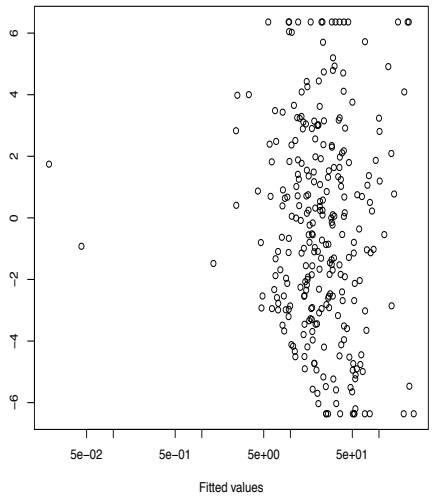


ZMNB

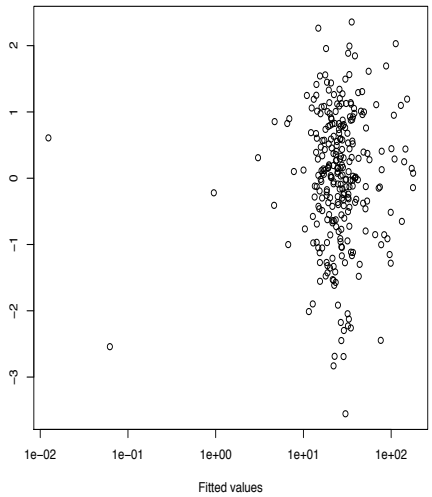


ZINB

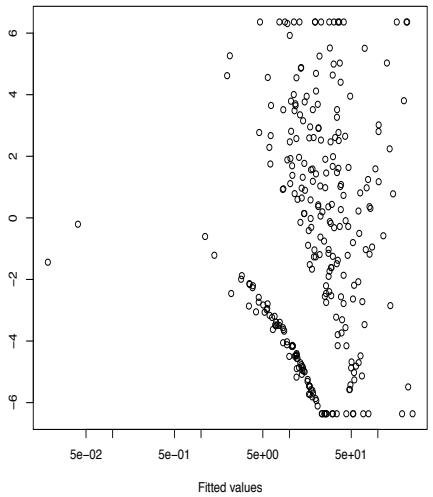
RQR vs Fitted Values for "Euba"



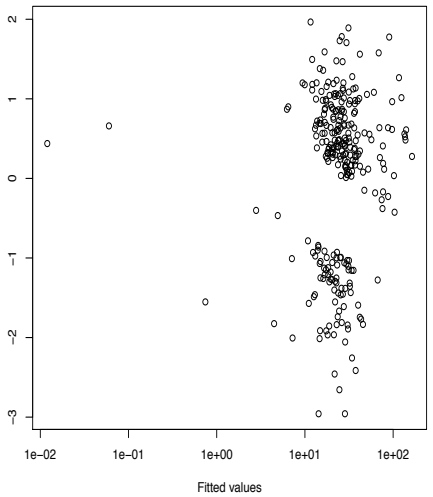
Poisson1



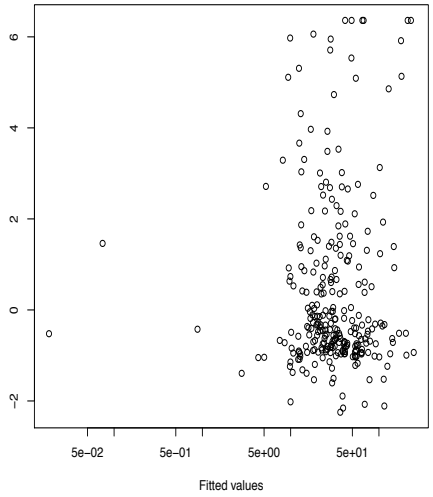
NB1



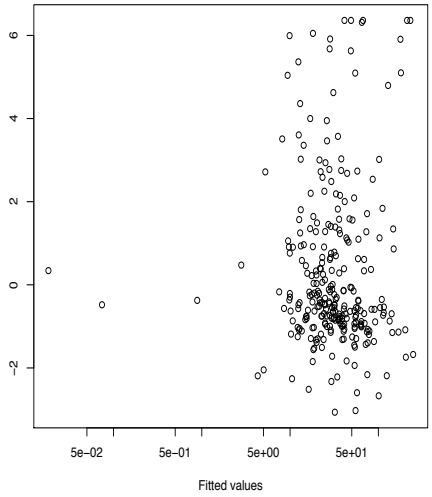
Poisson



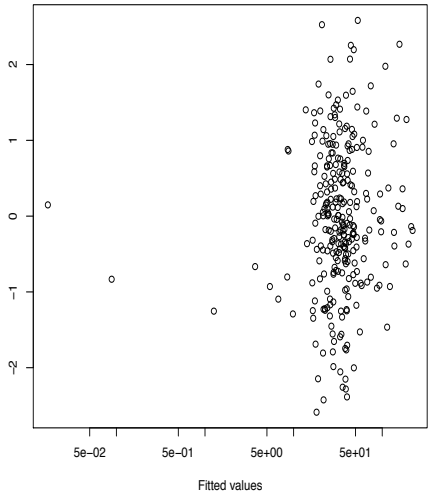
NB



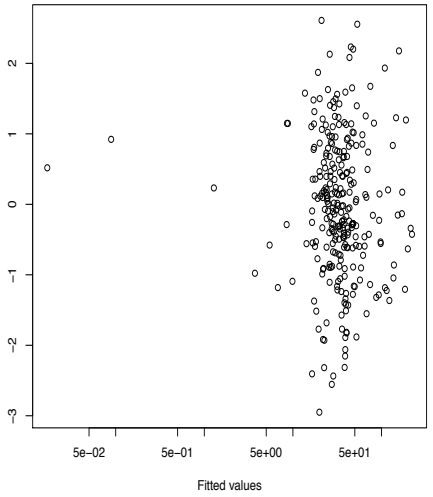
ZIP



ZMP

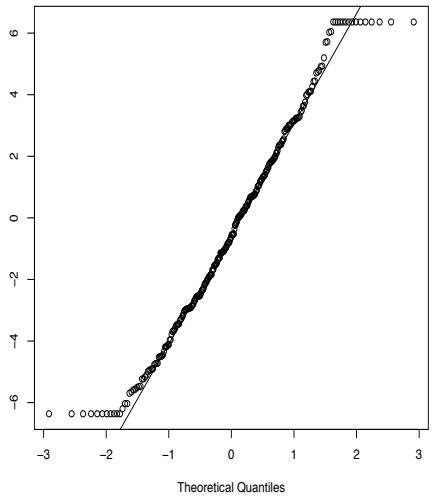


ZINB

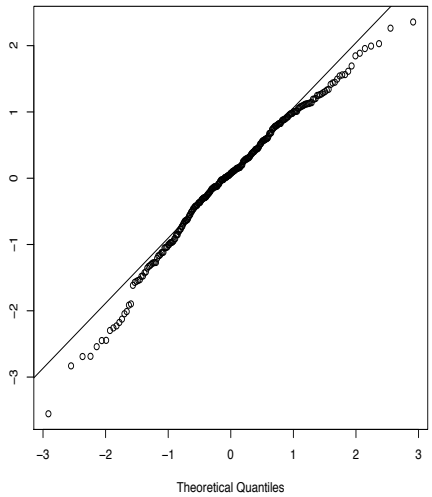


ZMNB

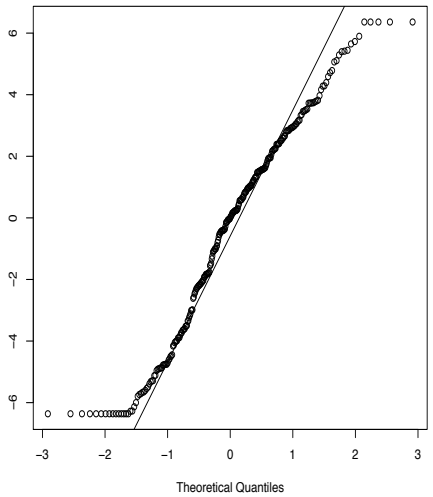
QQ-plot of RQR for "Euba"



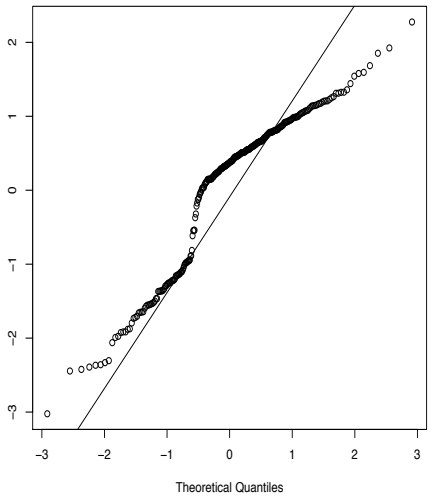
Poisson1



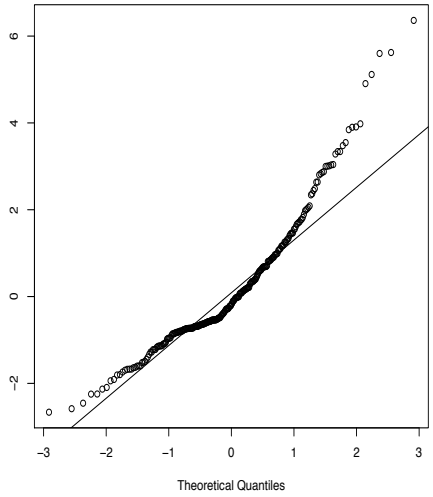
NB1



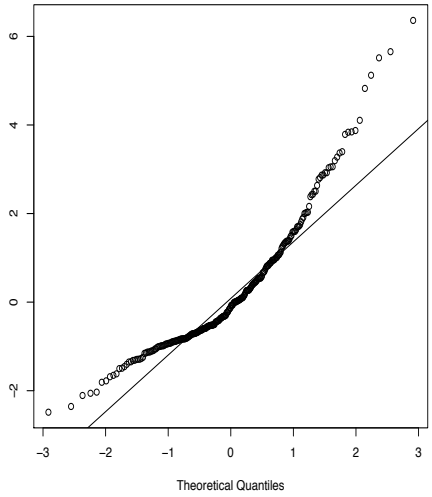
Poisson



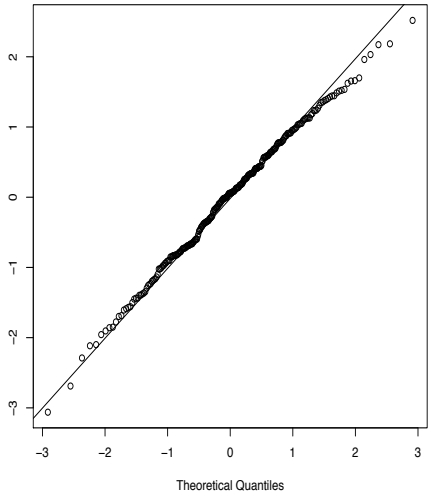
NB



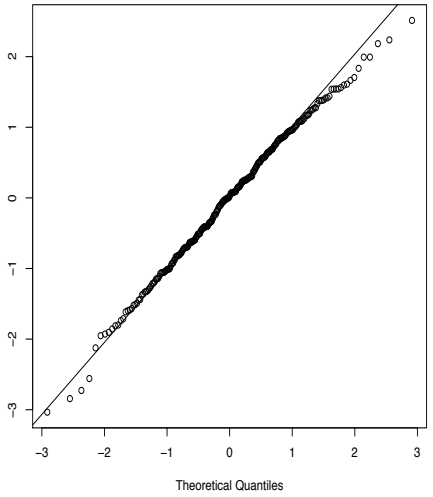
ZIP



ZMP



ZINB



ZMNB

Shapiro-Wilk P-values for all 14 Genera

Table 5: P-values for the Shapiro-Wilk test of Randomized quantile residuals for twin study OTU data sorted by ZMNB

Genus	ZMNB	ZINB	ZMP	ZIP	NB	Poisson	NB1	Poisson1
Bact	0.052	0.034	$< 10^{-19}$	$< 10^{-19}$	$< 10^{-16}$	$< 10^{-18}$	$< 10^{-8}$	$< 10^{-17}$
Lach..g	0.072	0.074	$< 10^{-16}$	$< 10^{-15}$	$< 10^{-3}$	$< 10^{-11}$	0.005	$< 10^{-4}$
Faec	0.083	0.107	$< 10^{-17}$	$< 10^{-18}$	$< 10^{-17}$	$< 10^{-15}$	$< 10^{-10}$	$< 10^{-13}$
Rumi	0.232	0.285	$< 10^{-19}$	$< 10^{-19}$	$< 10^{-6}$	$< 10^{-12}$	0.04	$< 10^{-5}$
Rumi.1	0.238	0.366	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-10}$	$< 10^{-11}$	$< 10^{-5}$	$< 10^{-10}$
Blau	0.251	0.104	$< 10^{-10}$	$< 10^{-10}$	0.087	$< 10^{-12}$	0.182	$< 10^{-12}$
Erys	0.344	0.258	$< 10^{-16}$	$< 10^{-17}$	$< 10^{-4}$	$< 10^{-7}$	0.314	$< 10^{-5}$
Alis	0.344	0.352	$< 10^{-16}$	$< 10^{-16}$	$< 10^{-9}$	$< 10^{-7}$	0.003	$< 10^{-6}$
Euba	0.461	0.539	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-10}$	$< 10^{-6}$	0.006	$< 10^{-4}$
Lach	0.521	0.358	$< 10^{-9}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-5}$	0.003	0.051
Oscil	0.535	0.606	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-9}$	$< 10^{-5}$	0.006	$< 10^{-4}$
Prev	0.605	0.269	$< 10^{-17}$	$< 10^{-17}$	$< 10^{-4}$	$< 10^{-12}$	0.002	$< 10^{-12}$
Rose	0.627	0.613	$< 10^{-13}$	$< 10^{-14}$	$< 10^{-6}$	$< 10^{-13}$	0.749	$< 10^{-13}$
Copr	0.752	0.721	$< 10^{-13}$	$< 10^{-14}$	$< 10^{-8}$	$< 10^{-6}$	0.245	$< 10^{-6}$

Conclusions and Discussions

- Our studies show that RQR performs very well for checking GLMM. RQRs are normally distributed under the true model. In GOF test, the type 1 error rates of RQR are close to the nominal level 0.05, and the statistical powers of RQR in rejecting wrong models are very good.
- We have applied RQR to assess models for a real human microbiome dataset at genus level and found that ZMNB and ZINB are good models for the dataset and other simpler models (such as NB and Poisson) are not adequate to describe the extraordinarily small and large OTU counts.
- We have developed generic functions for computing RQRs with fitting outputs of R package `glmmTMB`. They will be released in Wei Bai's M.Sc. thesis.