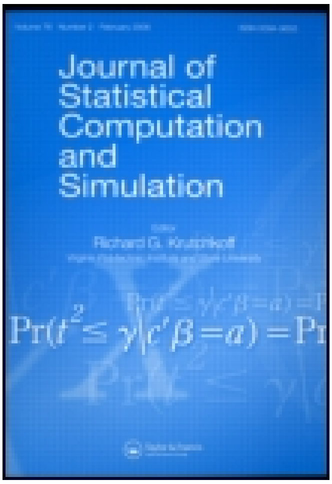

Journal of Statistical Computation and Simulation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gscs20

Parametric simultaneous robust inferences for regression coefficient under generalized linear models

Li-Chu Chien[a] & Tsung-Shan Tsou[bcd]

[a] Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan, Republic of China

[b] Institute of Statistics, National Central University, Jhongli, Taiwan, Republic of China

[c] Institute of Systems Biology and Bioinformatics, Center for Biotechnology and Biomedical Engineering, National Central University, Jhongli, Taiwan, Republic of China

[d] Cathay Medical Research Institute, Cathay General Hospital, Taipei, Taiwan, Republic of China

Published online: 15 Oct 2012.

To cite this article: Li-Chu Chien & Tsung-Shan Tsou (2014) Parametric simultaneous robust inferences for regression coefficient under generalized linear models, Journal of Statistical Computation and Simulation, 84:4, 850-867, DOI: 10.1080/00949655.2012.731409

To link to this article: http://dx.doi.org/10.1080/00949655.2012.731409

Journal of Statistical Computation and Simulation, 2014
Vol. 84, No. 4, 850–867, http://dx.doi.org/10.1080/00949655.2012.731409

Parametric simultaneous robust inferences for regression coefficient under generalized linear models

Li-Chu Chien[a]* and Tsung-Shan Tsou[b,c,d]

[a]*Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan, Republic of China;* [b]*Institute of Statistics, National Central University, Jhongli, Taiwan, Republic of China;* [c]*Institute of Systems Biology and Bioinformatics, Center for Biotechnology and Biomedical Engineering, National Central University, Jhongli, Taiwan, Republic of China;* [d]*Cathay Medical Research Institute, Cathay General Hospital, Taipei, Taiwan, Republic of China*

(*Received 4 October 2011; final version received 14 September 2012*)

In this article, the parametric robust regression approaches are proposed for making inferences about regression parameters in the setting of generalized linear models (GLMs). The proposed methods are able to test hypotheses on the regression coefficients in the misspecified GLMs. More specifically, it is demonstrated that with large samples, the normal and gamma regression models can be properly adjusted to become asymptotically valid for inferences about regression parameters under model misspecification. These adjusted regression models can provide the correct type I and II error probabilities and the correct coverage probability for continuous data, as long as the true underlying distributions have finite second moments.

Keywords: generalized linear models; robust normal regression; robust gamma regression

1. Introduction

Generalized linear models (GLMs) were introduced by Nelder and Wedderburn [1] as a unifying family of models for non-standard cross-sectional regression analysis with non-normal responses. The statistical analysis of such models is based on the asymptotic properties of the maximum likelihood estimator (MLE). Fahrmeir and Kaufmann [2] presented mild general conditions, which, respectively, assure weak or strong consistency or asymptotic normality of the MLE. More on this study can be found in [3]. More generally, Fahrmeir [4] dealt with the asymptotic behaviour of the quasi-MLE in misspecified GLMs.

Cantoni and Ronchetti [5] proposed a natural class of robust estimation techniques for GLMs. Their method is more reliable than the classical estimation procedures in providing the accurate statistical inference when the data include outlying points. Adimari and Ventura [6] also studied robust inference for GLMs. They derived a robust quasi-profile log-likelihood function that was obtained from an estimating function that defines the class of Mallows-type robust estimators considered by Cantoni and Ronchetti [5]. Li and Hsiao [7] suggested a method for consistently estimating GLMs with measurement errors without making any prior distributional assumption on

*Corresponding author. Email: lcchien@nctu.edu.tw

the measurement error or the latent variables. However, the robustness of their proposed method requires the knowledge of the probability distribution of latent variables. Sinha [8] developed a robust method for analysing GLMs with non-ignorable missing covariates. Recently, Bianco *et al.* [9] introduced a resistant procedure to test hypotheses on the regression parameter in GLMs with missing responses.

On the other hand, Heagerty and Kurland [10] evaluated the impact of model violations on the estimate of a regression coefficient in generalized linear mixed models (GLMMs). Jiang and Zhang [11] proposed robust methods to estimate parameters of interest in settings of GLMMs, in which only the conditional means of the responses given the random effects are specified. Yau and Kuk [12] proposed robust estimation procedures for GLMMs based on the notion of maximum quasi-likelihood and residual maximum quasi-likelihood. Sinha [13] developed a robust method for identifying and downweighting the outliers when estimating the parameters in the GLMMs. Sinha [14] further described a robust quasi-likelihood method for fitting the GLMMs to longitudinal data.

In addition, robust restricted maximum likelihood (robust REML) in mixed linear models are introduced by Richardson and Welsh [15] who made classical REML robust by bounding the influence of outlying observations on the estimate. Yun and Lee [16] discussed the robust estimation in mixed linear models with non-monotone missingness. Jacqmin-Gadda *et al.* [17] investigated the robustness of the MLE of fixed effects from a linear mixed model when the error terms are either correlated or non-Gaussian or of non-constant variance.

Royall and Tsou [18] advocated the robust likelihood function concept. They developed a technique that adjusts a working likelihood function, making it robust. The resulting adjusted robust likelihood function remains valid evidential representation of the parameter, even when the working model is incorrect. Motivated by the above results, Tsou [19] proposed a parametric robust way for comparing two population means and two population variances in misspecified models. Tsou and Cheng [20] applied the robust likelihood techniques to analyse contaminated data in regression settings. Tsou [21] further extended the robust likelihood concept to analyse count data. In this article, the robust likelihood techniques are used to make inferences about regression parameters in the GLM setting.

This article is organized as follows. Section 2 contains a brief review of the idea of robust likelihood functions introduced by Royall and Tsou [18]. The robust normal regression (RNR) and robust gamma regression (RGR) are briefly introduced in Section 3. Section 4 presents a simulation study which shows the advantage of the RNR and RGR models with respect to (w.r.t.) the ordinary normal and gamma regression models. Section 5 concludes with a brief discussion. Some technical background material from the previous sections is deferred into the appendix.

2. Robust likelihood functions

Suppose that $Y_1, Y_2, \dots, Y_N$ is a sequence of independent random variables. On the basis of *a priori* knowledge or convenience, we postulate a working model for the probability distributions of Y_i's, $\{f_i = f_i(\bullet; \boldsymbol{\psi}) = f(\bullet; \eta_i(\boldsymbol{\psi})), i = 1, 2, \dots, N, \boldsymbol{\psi} \in \Psi\}$, where $\boldsymbol{\psi}$ is a fixed-dimensional vector of unknown parameters. For example, under normal regression settings, $\eta_i(\boldsymbol{\psi}) = (\boldsymbol{x}_i^t\boldsymbol{\gamma}, \sigma^2)$, $\boldsymbol{\psi} = (\boldsymbol{\gamma}^t, \sigma^2)^t$ and $f_i = f_i(y_i; \boldsymbol{\psi}) = \exp\{-(y_i - \boldsymbol{x}_i^t\boldsymbol{\gamma})^2/2\sigma^2\}/\sqrt{2\pi}\sigma$. Here $\boldsymbol{x}_i$ represents the p characteristics that are specific to y_i, and $\boldsymbol{\gamma}$ represents the p regression coefficients that describe how $\boldsymbol{x}_i$ affects the expected value of Y_i. Note that this model is a collection of probability distributions, each of which is identified by a unique value of $\boldsymbol{\psi}$.

Now partition $\boldsymbol{\psi}$ as $\boldsymbol{\psi}^t = (\boldsymbol{\theta}^t, \boldsymbol{\varphi}^t)$, where $\boldsymbol{\theta}$ is the p-vector of parameters of interest and $\boldsymbol{\varphi}$ is the remaining fixed-dimensional nuisance parameters. Let $\boldsymbol{\theta}_0$ and $\boldsymbol{\varphi}_0$ denote the limiting values of the MLEs, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$, based on the working model $f = (f_1, f_2, \ldots, f_N)$, when the Y_i's are actually generated from the family $\{h_i = h(\bullet; \tau_i(\boldsymbol{\theta}, \boldsymbol{\lambda})), i = 1, 2, \ldots, N\}$, where $\boldsymbol{\lambda}$ is the nuisance parameter vector under $h = (h_1, h_2, \ldots, h_N)$. Now suppose that the parameters of inference under the working model f, namely $\boldsymbol{\theta}$, remain the parameters of interest under h, so that $\boldsymbol{\theta}_0$ has the same interpretation of the true values of the parameters of interest. This result is what Royall and Tsou [18] referred to as the first condition of robustness (FCR). This condition is crucial for the working model to be adjustable for valid inferences. Note that White [22] showed that, more often than not, the FCR is not satisfied once $f \neq h$.

Write $l_{\boldsymbol{\theta}}$ and $l_{\boldsymbol{\varphi}}$ for the first derivatives of the log-likelihood function $l(\boldsymbol{\theta}, \boldsymbol{\varphi})$ w.r.t. $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$, respectively, whose derivatives w.r.t. $\boldsymbol{\varphi}$ are correspondingly denoted by $l_{\boldsymbol{\theta\varphi}}$ and $l_{\boldsymbol{\varphi\varphi}}$. Now, let $\boldsymbol{I}_{h\boldsymbol{\theta\varphi}}$ and $\boldsymbol{I}_{h\boldsymbol{\varphi\varphi}}$ be the limiting values of $-l_{\boldsymbol{\theta\varphi}}/N$ and $-l_{\boldsymbol{\varphi\varphi}}/N$, respectively, under h and the limiting values of $-l_{\boldsymbol{\theta\theta}}/N$ and $-l_{\boldsymbol{\varphi\theta}}/N$, under h, are denoted by $\boldsymbol{I}_{h\boldsymbol{\theta\theta}}$ and $\boldsymbol{I}_{h\boldsymbol{\varphi\theta}}$, respectively. Note that these limiting values are all evaluated at $\boldsymbol{\theta}_0$ and $\boldsymbol{\varphi}_0$.

Now define the following two $p \times p$ matrices:

$$\boldsymbol{A} = \boldsymbol{I}_{h\boldsymbol{\theta\theta}} - \boldsymbol{I}_{h\boldsymbol{\theta\varphi}}\boldsymbol{I}_{h\boldsymbol{\varphi\varphi}}^{-1}\boldsymbol{I}_{h\boldsymbol{\varphi\theta}} \tag{1}$$

and

$$\boldsymbol{B} = \boldsymbol{V}_{h\boldsymbol{\theta\theta}} - \boldsymbol{I}_{h\boldsymbol{\theta\varphi}}\boldsymbol{I}_{h\boldsymbol{\varphi\varphi}}^{-1}\boldsymbol{V}_{h\boldsymbol{\varphi\theta}} - \boldsymbol{V}_{h\boldsymbol{\theta\varphi}}\boldsymbol{I}_{h\boldsymbol{\varphi\varphi}}^{-1}\boldsymbol{I}_{h\boldsymbol{\varphi\theta}} + \boldsymbol{I}_{h\boldsymbol{\theta\varphi}}\boldsymbol{I}_{h\boldsymbol{\varphi\varphi}}^{-1}\boldsymbol{V}_{h\boldsymbol{\varphi\varphi}}\boldsymbol{I}_{h\boldsymbol{\varphi\varphi}}^{-1}\boldsymbol{I}_{h\boldsymbol{\varphi\theta}}. \tag{2}$$

Here $\boldsymbol{V}_{h\boldsymbol{\theta\theta}} = \lim_{N\to\infty} E_h[l_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \boldsymbol{\varphi}_0)l_{\boldsymbol{\theta}}^t(\boldsymbol{\theta}_0, \boldsymbol{\varphi}_0)/N]$, $\boldsymbol{V}_{h\boldsymbol{\theta\varphi}} = \lim_{N\to\infty} E_h[l_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \boldsymbol{\varphi}_0)l_{\boldsymbol{\varphi}}^t(\boldsymbol{\theta}_0, \boldsymbol{\varphi}_0)/N]$ and $\boldsymbol{V}_{h\boldsymbol{\varphi\varphi}} = \lim_{N\to\infty} E_h[l_{\boldsymbol{\varphi}}(\boldsymbol{\theta}_0, \boldsymbol{\varphi}_0)l_{\boldsymbol{\varphi}}^t(\boldsymbol{\theta}_0, \boldsymbol{\varphi}_0)/N]$, where E_h stands for the expectation evaluated under h.

Let $\hat{\boldsymbol{\theta}}$ be the MLE of $\boldsymbol{\theta}$ and $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{B}}$ be the empirical versions of $\boldsymbol{A}$ and $\boldsymbol{B}$. A direct application of Taylor's expansion shows that the adjusted Wald statistic $N(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^t\hat{\boldsymbol{A}}\hat{\boldsymbol{B}}^{-1}\hat{\boldsymbol{A}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ has an asymptotic χ_p^2 distribution for general $h_i, i = 1, 2, \ldots, N$, that have finite second moments. Here χ_p^2 is denoted as a chi-squared distribution with p degrees of freedom. Another asymptotically equivalent counterpart, the adjusted score statistic $N^{-1}\{l_{\boldsymbol{\theta}}^t(\boldsymbol{\theta}_0, \hat{\boldsymbol{\varphi}}(\boldsymbol{\theta}_0))\}\hat{\boldsymbol{B}}^{-1}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\varphi}}(\boldsymbol{\theta}_0))\{l_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0, \hat{\boldsymbol{\varphi}}(\boldsymbol{\theta}_0))\}$, where $\hat{\boldsymbol{\varphi}}(\boldsymbol{\theta}_0)$ and $\hat{\boldsymbol{B}}(\boldsymbol{\theta}_0, \boldsymbol{\varphi}(\boldsymbol{\theta}_0))$ are the constrained MLEs of $\boldsymbol{\varphi}$ and $\boldsymbol{B}$ given $\boldsymbol{\theta}_0$, respectively, has the same limiting χ_p^2 distribution even if the working model assumptions fail.

3. Robust regression models

Consider a set of observations $y_1, y_2, \ldots, y_N$ corresponding to N independent not identically distributed random variables $Y_1, Y_2, \ldots, Y_N$. Under GLMs, the mean response, μ_i, depends on the p covariates $(x_{i,0}, x_{i,1}, \ldots, x_{i,p-1}) = \boldsymbol{x}_i^t$, by $\mu_i = g(\eta_i)$, where $\eta_i = \boldsymbol{x}_i^t\boldsymbol{\gamma} = \gamma_0 x_{i,0} + \gamma_1 x_{i,1} + \cdots + \gamma_{p-1}x_{i,p-1}$ is a linear predicator with the p regression coefficients $(\gamma_0, \gamma_1, \ldots, \gamma_{p-1}) = \boldsymbol{\gamma}^t$, and $g(\bullet)$ is a monotonic and differentiable response function.

3.1. *Robust normal regression*

Under a normal working model, the log-likelihood function for the ith observation y_i is

$$l_i = -\frac{1}{2}\log\sigma^2 - \frac{1}{2}\log 2\pi - \frac{(y_i - \mu_i)^2}{2\sigma^2}.$$

The log-likelihood equation for γ_{j-1} is

$$\frac{1}{\sigma^2}\sum_{i=1}^{N}\mu_i'(y_i-\mu_i)x_{i,j-1}=0,\quad j=1,2,\ldots,p, \tag{3}$$

where μ_i' is the first derivative of μ_i w.r.t. η_i. The solutions of Equation (3) are the maximum quasi-likelihood (MQL) estimators [23–25] or M-estimators [26], when the observations $y_1, y_2, \ldots, y_N$ are not necessarily from normal distributions. McCullagh [27] showed that, under mild regularity conditions, the consistency of the MQL estimates under model misspecification depends only on the correct specification of the regression. In other words, the normal working model provides the consistent estimates of regression parameters under incorrectly specified models. Thus, the FCR is fulfilled, so that the normal working model can be properly adjusted to become asymptotically legitimate for regression parameters of interest under model misspecification.

Without loss of generality, let $\gamma_{p-w}, \gamma_{p-w+1}, \ldots, \gamma_{p-2}, \gamma_{p-1}$ be the w parameters of interest and let $(\gamma_{p-1}, \gamma_{p-2}, \ldots, \gamma_{p-w+1}, \gamma_{p-w})$ be denoted by $(\beta_1, \beta_2, \ldots, \beta_{w-1}, \beta_w) = \boldsymbol{\beta}^t$ for notational convenience. Let $\mu_{i,0}$ and $\mu_{i,0}'$ be, respectively, the true values of μ_i and μ_i'. Let $\mathrm{Var}_h(Y_i), i = 1, 2, \ldots, N$, be the true variances of $Y_i, i = 1, 2, \ldots, N$. Let $\boldsymbol{Z} = (z_0, z_1, \ldots, z_{p-1})$ be the $N \times p$ design matrix, so that $\boldsymbol{Z}^t = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)$.

After lengthy derivations, it shows that the $(u, v), u, v = 1, 2, \ldots, w$, elements of the $w \times w$ adjusting matrices $\boldsymbol{A}_n$ and $\boldsymbol{B}_n$ of $\boldsymbol{A}_n\boldsymbol{B}_n^{-1}\boldsymbol{A}_n$ that make the normal regression model robust can be written in the forms (for details, see the appendix):

$$\boldsymbol{A}_{n(uv)}=\lim_{N\to\infty}\frac{1}{N\sigma_0^2}\sum_{i=1}^{N}(\mu_{i,0}')^2\left(x_{i,p-u}-\sum_{j=1}^{p-w}\frac{|\boldsymbol{\Delta}_{nj(u)}|}{|\boldsymbol{\Delta}_n|}x_{i,j-1}\right)\left(x_{i,p-v}-\sum_{j=1}^{p-w}\frac{|\boldsymbol{\Delta}_{nj(v)}|}{|\boldsymbol{\Delta}_n|}x_{i,j-1}\right)$$

and

$$\boldsymbol{B}_{n(uv)}=\lim_{N\to\infty}\frac{1}{N\sigma_0^4}\sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu_{i,0}')^2\left(x_{i,p-u}-\sum_{j=1}^{p-w}\frac{|\boldsymbol{\Delta}_{nj(u)}|}{|\boldsymbol{\Delta}_n|}x_{i,j-1}\right)\times\left(x_{i,p-v}-\sum_{j=1}^{p-w}\frac{|\boldsymbol{\Delta}_{nj(v)}|}{|\boldsymbol{\Delta}_n|}x_{i,j-1}\right).$$

Here $|\boldsymbol{\Delta}_n|$ represents the determinant of the matrix $\boldsymbol{\Delta}_n$, where $\boldsymbol{\Delta}_n = \boldsymbol{W}^t\boldsymbol{V}_n^{-1}\boldsymbol{W}$ with $\boldsymbol{W} = (z_0, \ldots, z_{j-2}, z_{j-1}, z_j, \ldots, z_{p-w-1})$ and $\boldsymbol{V}_n = \mathrm{diag}((1/\mu_{1,0}')^2, (1/\mu_{2,0}')^2, \ldots, (1/\mu_{N,0}')^2)$ being a diagonal matrix of order N. On the other hand, $\boldsymbol{\Delta}_{nj(u)} = \boldsymbol{W}^t\boldsymbol{V}_n^{-1}\boldsymbol{W}_{j(u)}$ and $\boldsymbol{\Delta}_{nj(v)} = \boldsymbol{W}^t\boldsymbol{V}_n^{-1}\boldsymbol{W}_{j(v)}$ with $\boldsymbol{W}_{j(u)} = (z_0, \ldots, z_{j-2}, z_{p-u}, z_j, \ldots, z_{p-w-1})$ and $\boldsymbol{W}_{j(v)} = (z_0, \ldots, z_{j-2}, z_{p-v}, z_j, \ldots, z_{p-w-1})$ derived by the jth column z_{j-1} of $\boldsymbol{W}$ replaced by z_{p-u} and z_{p-v}, respectively. Here σ_0^2 is the limit of the MLE of σ^2, $\hat{\sigma}^2$, that has the same interpretation of the limit of $\sum_{i=1}^{N}\mathrm{Var}_h(Y_i)/N$. Note that the interpretation of σ_0^2 depends on h and is, therefore, unknown.

In the special case with all the regression coefficients of interest, let $(\gamma_{p-1}, \gamma_{p-2}, \ldots, \gamma_1, \gamma_0) = (\beta_1, \beta_2, \ldots, \beta_{p-1}, \beta_p) = \boldsymbol{\beta}^t$. Then, the adjusting matrices $\boldsymbol{A}_n$ and $\boldsymbol{B}_n$ can be simplified as follows:

$$\boldsymbol{A}_{n(uv)}=\lim_{N\to\infty}\frac{1}{N\sigma_0^2}\sum_{i=1}^{N}(\mu_{i,0}')^2(x_{i,p-u})(x_{i,p-v})$$

and

$$\boldsymbol{B}_{n(uv)}=\lim_{N\to\infty}\frac{1}{N\sigma_0^4}\sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu_{i,0}')^2(x_{i,p-u})(x_{i,p-v}).$$

In applications, consistent estimates $\hat{\boldsymbol{A}}_n$ and $\hat{\boldsymbol{B}}_n$ of $\boldsymbol{A}_n$ and $\boldsymbol{B}_n$ can be obtained by $\mathrm{Var}_h(Y_i)$ replaced by $(y_i - \hat{\mu}_i)^2$ with $\hat{\mu}$ being the MLE of μ and other unknown quantities replaced by their respective empirical versions.

Let $\boldsymbol{\beta}_0$ be the true value of $\boldsymbol{\beta}$ and consider the null hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. Let $\hat{\boldsymbol{\beta}}$ be the MLE of $\boldsymbol{\beta}$ based on the normal working model and let $\hat{\boldsymbol{\varphi}}(\boldsymbol{\beta}_0)$ and $\hat{\boldsymbol{B}}_n(\boldsymbol{\beta}_0, \hat{\boldsymbol{\varphi}}(\boldsymbol{\beta}_0))$ be the restricted MLEs of $\boldsymbol{\varphi}$ and $\boldsymbol{B}_n$ given $\boldsymbol{\beta}_0$. Here the vector of the nuisance parameters, $\boldsymbol{\varphi}$, contains the scale parameter σ^2 and some regression coefficients that are not to be tested. Under H_0, the adjusted Wald statistic $N(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^t \hat{\boldsymbol{A}}_n \hat{\boldsymbol{B}}_n^{-1} \hat{\boldsymbol{A}}_n (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ and the adjusted score statistic $N^{-1}\{l_{\boldsymbol{\beta}}^t(\boldsymbol{\beta}_0, \boldsymbol{\varphi}(\boldsymbol{\beta}_0))\}\hat{\boldsymbol{B}}_n^{-1}(\boldsymbol{\beta}_0, \hat{\boldsymbol{\varphi}}(\boldsymbol{\beta}_0))\{l_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \hat{\boldsymbol{\varphi}}(\boldsymbol{\beta}_0))\}$ are asymptotically equivalent and have an asymptotic χ_w^2 distribution as long as the second moments of the true underlying distributions exist. Note that $\hat{\boldsymbol{A}}_n \hat{\boldsymbol{B}}_n^{-1} \hat{\boldsymbol{A}}_n$ is free of σ^2. Thus, with large samples, the effect of σ^2 is actually removed. Hence, σ^2 can be treated known, *a priori*, as any arbitrary positive value.

3.2. *Robust gamma regression*

Under a gamma working model, the log-likelihood function for the ith observation y_i is

$$l_i = r \log r - r \log \mu_i + (r-1) \log y_i - r\mu_i^{-1} y_i - \log \Gamma(r).$$

The score functions

$$r \sum_{i=1}^{N} \frac{\mu_i'}{\mu_i} \left(\frac{y_i - \mu_i}{\mu_i} \right) x_{i,j-1}, \quad j = 1, 2, \ldots, p,$$

have zero expectation as long as $\mu_i, i = 1, 2, \ldots, N$, are correctly specified. Hence, the regression parameters of interest can be consistently estimated by the gamma working model, whatever h is. Thus, the FCR is satisfied.

Calculations parallel to $\boldsymbol{A}_n$ and $\boldsymbol{B}_n$ show that the $(u, v), u, v = 1, 2, \ldots, w$, components of the adjusting matrices $\boldsymbol{A}_g$ and $\boldsymbol{B}_g$ under the gamma working model are of the forms (for details, see the appendix):

$$\boldsymbol{A}_{g(uv)} = \lim_{N\to\infty} \frac{r_0}{N} \sum_{i=1}^{N} \left(\frac{\mu_{i,0}'}{\mu_{i,0}} \right)^2 \left(x_{i,p-u} - \sum_{j=1}^{p-w} \frac{|\boldsymbol{\Delta}_{gj(u)}|}{|\boldsymbol{\Delta}_g|} x_{i,j-1} \right) \left(x_{i,p-v} - \sum_{j=1}^{p-w} \frac{|\boldsymbol{\Delta}_{gj(v)}|}{|\boldsymbol{\Delta}_g|} x_{i,j-1} \right)$$

and

$$\boldsymbol{B}_{g(uv)} = \lim_{N\to\infty} \frac{r_0^2}{N} \sum_{i=1}^{N} \frac{\mathrm{Var}_h(Y_i)}{\mu_{i,0}^2} \left(\frac{\mu_{i,0}'}{\mu_{i,0}} \right)^2 \left(x_{i,p-u} - \sum_{j=1}^{p-w} \frac{|\boldsymbol{\Delta}_{gj(u)}|}{|\boldsymbol{\Delta}_g|} x_{i,j-1} \right) \times \left(x_{i,p-v} - \sum_{j=1}^{p-w} \frac{|\boldsymbol{\Delta}_{gj(v)}|}{|\boldsymbol{\Delta}_g|} x_{i,j-1} \right),$$

where $\boldsymbol{\Delta}_g = \boldsymbol{W}^t \boldsymbol{V}_g^{-1} \boldsymbol{W}$, $\boldsymbol{\Delta}_{gj(u)} = \boldsymbol{W}^t \boldsymbol{V}_g^{-1} \boldsymbol{W}_{j(u)}$ and $\boldsymbol{\Delta}_{gj(v)} = \boldsymbol{W}^t \boldsymbol{V}_g^{-1} \boldsymbol{W}_{j(v)}$ with $\boldsymbol{V}_g = \mathrm{diag}((\mu_{1,0}/\mu_{1,0}')^2, (\mu_{2,0}/\mu_{2,0}')^2, \ldots, (\mu_{N,0}/\mu_{N,0}')^2)$. Here r_0 is the limit of the MLE of r, whose interpretation depends on h and is, therefore, unknown.

In the special case with all the regression parameters of interest, $\boldsymbol{A}_g$ and $\boldsymbol{B}_g$ reduce to

$$\boldsymbol{A}_{g(uv)} = \lim_{N\to\infty} \frac{r_0}{N} \sum_{i=1}^{N} \left(\frac{\mu'_{i,0}}{\mu_{i,0}}\right)^2 (x_{i,p-u})(x_{i,p-v})$$

and

$$\boldsymbol{B}_{g(uv)} = \lim_{N\to\infty} \frac{r_0^2}{N} \sum_{i=1}^{N} \frac{\mathrm{Var}_h(Y_i)}{\mu_{i,0}^2} \left(\frac{\mu'_{i,0}}{\mu_{i,0}}\right)^2 (x_{i,p-u})(x_{i,p-v}).$$

In application, consistent estimates $\hat{\boldsymbol{A}}_g$ and $\hat{\boldsymbol{B}}_g$ of $\boldsymbol{A}_g$ and $\boldsymbol{B}_g$ can be derived by replacing the unknown components in $\boldsymbol{A}_g$ and $\boldsymbol{B}_g$ by their respective empirical analogues, just as we dealt with $\hat{\boldsymbol{A}}_n$ and $\hat{\boldsymbol{B}}_n$. Note that $\hat{\boldsymbol{A}}_g\hat{\boldsymbol{B}}_g^{-1}\hat{\boldsymbol{A}}_g$ is free of r, so that with large samples, the effect of r is completely eliminated. Therefore, r can be treated known, *a priori*, in the beginning as any positive value. The resulting adjusted Wald statistic $N(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^t\hat{\boldsymbol{A}}_g\hat{\boldsymbol{B}}_g^{-1}\hat{\boldsymbol{A}}_g(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ and the resulting adjusted score statistic $N^{-1}\{l^t_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \hat{\boldsymbol{\varphi}}(\boldsymbol{\beta}_0))\}\hat{\boldsymbol{B}}_g^{-1}(\boldsymbol{\beta}_0, \hat{\boldsymbol{\varphi}}(\boldsymbol{\beta}_0))\{l_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0, \hat{\boldsymbol{\varphi}}(\boldsymbol{\beta}_0))\}$, under $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, are asymptotically distributed as χ^2_w for general h with the finite second moments. Note that here all MLEs are derived under the gamma working model.

4. Simulation studies

To investigate the performance of the RNR and RGR models in the finite sample situation, simulation studies are conducted using $N = 450$, 900 and 1350 replicated samples, respectively, generated from the three regression models

$$\text{Model 1:} \quad \mu_i = \exp(\eta_i),$$
$$\text{Model 2:} \quad \mu_i = (2.5\eta_i + 2/3)^3,$$
$$\text{Model 3:} \quad \mu_i = \eta_i^2$$

with the linear predicator η_i given by

$$\eta_i = x_{i,0} + \gamma_1 x_{i,1} + \gamma_2 x_{i,2} \quad \text{for } i = 1, 2, \ldots, N,$$

where the values of $x_{i,0}, i = 1, 2, \ldots, N$, are set by 1 and the values of $x_{i,j}, i = 1, 2, \ldots, N, j = 1, 2$, are independently generated from a uniform distribution between 0 and 1. Here regression coefficients γ_1 and γ_2 are considered as the parameters of interest. For simplicity, let $\boldsymbol{\gamma}^t = (\gamma_1, \gamma_2)$ and $\boldsymbol{\gamma}_0^t = (1.0, 1.0)$. We test the null hypothesis $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}_0$ and the two alternative hypotheses $H_A : \boldsymbol{\gamma}^t = (0.4, 1.0)$ and $\boldsymbol{\gamma}^t = (0.7, 1.3)$, respectively.

Simulated data sets are generated from three sources including the Weibull, inverse Gaussian and chi-squared distributions, respectively. A Weibull distribution with the shape parameter λ and the scale parameter k, $W(k, \lambda)$, has a simple relationship between the second central moment and the first moment, that is, $\mathrm{Var}(Y) = a\mu^2$, where $a > 0$ is a function of the shape parameter λ. For example, when $\lambda = 1$ and $k = \mu$, $\mathrm{Var}(Y) = \mu^2$. Similarly, an inverse Gaussian distribution with the mean μ and the shape parameter λ, $\mathrm{IG}(\mu, \lambda)$, has a variance proportional to the cubic of its mean value, that is, $\mathrm{Var}(Y) = \mu^3/\lambda$. On the other hand, a non-central chi-squared distribution with ν degrees of freedom and a non-centrality parameter $\mu - \nu > 0$, $\chi^2_\nu(\mu - \nu)$, has a mean value of μ and a variance of $2(2\mu - \nu)$, so that $\chi^2_\nu(\mu - \nu)$ has a variance roughly proportional to its mean.

To demonstrate the robustness characters of the adjusted Wald and score statistics under the normal and gamma working models, in our simulations, the observations, $y_i, i = 1, 2, \ldots, N$, are sampled in the following way. First, the first $0.3N$ observations, $y_i, i = 1, 2, \ldots, 0.3N$, are independently generated from the Weibull distributions, $W(\mu_i, 1)$, with the shape parameter of 1 and the scale parameter of $\mu_i, i = 1, 2, \ldots, 0.3N$, respectively. Then, the next $0.3N$ observations, $y_i, i = 0.3N + 1, 0.3N + 2, \ldots, 0.6N$, are independently generated from the inverse Gaussian distributions, $\mathrm{IG}(\mu_i, 100)$, with the shape parameter of 100 and the mean value of $\mu_i, i = 0.3N + 1, 0.3N + 2, \ldots, 0.6N$, respectively. Finally, the rest of the $0.3N$ observations, $y_i, i = 0.6N + 1, 0.6N + 2, \ldots, N$, are independently generated from the non-central chi-squared distributions, $\chi_1^2(\mu_i - 1)$, with one degree of freedom and the non-centrality parameter of $\mu_i - 1, i = 0.6N + 1, 0.6N + 2, \ldots, N$, respectively.

Three additional test statistics are also included for contrast. They are the maximum likelihood ratio test statistic, the Wald test statistic and the score test statistic, respectively. The maximum likelihood ratio test statistic for testing the null hypothesis $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}_0$ is defined by

$$Q_{\mathrm{L}} = 2\{l(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\varphi}}) - l(\boldsymbol{\gamma}_0, \hat{\boldsymbol{\varphi}}(\boldsymbol{\gamma}_0))\}.$$

Then, its two asymptotically equivalent test statistics, the Wald test statistic and the score test statistic, are defined by

$$Q_{\mathrm{W}} = N(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)^t \hat{\boldsymbol{A}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$$

and

$$Q_{\mathrm{S}} = N^{-1}\{l_{\boldsymbol{\gamma}}^t(\boldsymbol{\gamma}_0, \hat{\boldsymbol{\varphi}}(\boldsymbol{\gamma}_0))\}\hat{\boldsymbol{A}}^{-1}(\boldsymbol{\gamma}_0, \hat{\boldsymbol{\varphi}}(\boldsymbol{\gamma}_0))\{l_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}_0, \hat{\boldsymbol{\varphi}}(\boldsymbol{\gamma}_0))\},$$

respectively. Here all notation definitions are given as in the previous sections. Each of the test statistics $Q_{\mathrm{L}}, Q_{\mathrm{W}}$ and Q_{S}, under the null hypothesis $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}_0$, has an asymptotic chi-squared distribution with degrees of freedom equal to the dimension of $\boldsymbol{\gamma}$. Thus, in our simulations, each of the maximum likelihood ratio test statistic, the Wald test statistic and the score test statistic rejects H_0, when each of the test statistics $Q_{\mathrm{L}}, Q_{\mathrm{W}}$ and Q_{S} exceeds the critical value of $\chi^2_{2,0.95}$, where $\chi^2_{2,0.95}$ represents the 95th quantile of the chi-squared distribution χ^2_2. More discussions about the test statistics $Q_{\mathrm{L}}, Q_{\mathrm{W}}$ and Q_{S} can be found in [28, Section 9.3].

The simulation performance are carried out for 3000 simulation runs with the $x_{i,1}$'s and $x_{i,2}$'s being regenerated after every 50 simulation runs. The empirical type I error probabilities based on the adjusted Wald statistic, the adjusted score statistic, the maximum likelihood ratio test statistic, the Wald test statistic and the score test statistic are labelled as $\mathrm{AW}\alpha, \mathrm{AS}\alpha, \mathrm{L}\alpha, \mathrm{W}\alpha$ and $\mathrm{S}\alpha$, respectively. On the other hand, AWcp, AScp, Lcp, Wcp and Scp symbolize the coverage probabilities of the nominal 95% confidence interval constructed using the adjusted Wald statistic, the adjusted score statistic, the maximum likelihood ratio test statistic, the Wald test statistic and the score test statistic, respectively. The empirical type I error probability is computed as the proportion of rejections of the null hypothesis $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}_0$ at the nominal 5% significance level, when the data are actually generated from H_0. On the other hand, when the data are sampled from the alternative hypothesis H_{A}, the empirical type I error probability exhibits the power of the test.

Results from the adjusted Wald statistic, the adjusted score statistic, the maximum likelihood ratio test statistic, the Wald test statistic and the score test statistic based on the normal and gamma working models are tabulated in the tables below. The average of the 3000 $\hat{\boldsymbol{\gamma}}$ values and their sample covariance matrix are termed as mean$(\hat{\boldsymbol{\gamma}})$ and S^2, respectively. In order to contrast the differences between the covariance matrix estimates based on the adjusted and unadjusted test statistics, the average of the unadjusted covariance matrix estimate of $\hat{\boldsymbol{\gamma}}$, namely $\hat{\boldsymbol{A}}^{-1}/N$, denoted

Table 1. Model 1: $\mu_i = \exp(\eta_i), i = 1, 2, \ldots, N$.

	Working model	mean($\hat{\boldsymbol{\gamma}}$)	S^2	$\mathrm{Var}_{\mathrm{A}}(\hat{\boldsymbol{\gamma}})$	$\mathrm{Var}(\hat{\boldsymbol{\gamma}})$	AWα	AWcp	ASα	AScp	Lα	Lcp	Wα	Wcp	Sα	Scp
$N = 450$															
$H_0 : \boldsymbol{\gamma} = \begin{pmatrix}1.0\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}1.0030\\1.0024\end{bmatrix}$	$\begin{bmatrix}0.0223 & 0.0054\\ & 0.0242\end{bmatrix}$	$\begin{bmatrix}0.0219 & 0.0040\\ & 0.0216\end{bmatrix}$	$\begin{bmatrix}0.0161 & 0.0001\\ & 0.0161\end{bmatrix}$	0.0643	0.9357	0.0587	0.9413	0.1190	0.8810	0.1223	0.8777	0.1157	0.8843
	Gamma	$\begin{bmatrix}0.9992\\0.9998\end{bmatrix}$	$\begin{bmatrix}0.0143 & 0.0005\\ & 0.0150\end{bmatrix}$	$\begin{bmatrix}0.0145 & 0.0004\\ & 0.0143\end{bmatrix}$	$\begin{bmatrix}0.0165 & 0.0001\\ & 0.0165\end{bmatrix}$	0.0597	0.9403	0.0407	0.9530	0.0307	0.9693	0.0290	0.9710	0.0320	0.9680
$H_{\mathrm{A}} : \boldsymbol{\gamma} = \begin{pmatrix}0.4\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}0.3980\\1.0038\end{bmatrix}$	$\begin{bmatrix}0.0198 & 0.0018\\ & 0.0203\end{bmatrix}$	$\begin{bmatrix}0.0193 & 0.0014\\ & 0.0194\end{bmatrix}$	$\begin{bmatrix}0.0152 & 0.0001\\ & 0.0177\end{bmatrix}$	0.9753	0.9377	0.9673	0.9413	0.9860	0.9133	0.9883	0.9137	0.9827	0.9187
	Gamma	$\begin{bmatrix}0.3980\\1.0016\end{bmatrix}$	$\begin{bmatrix}0.0154 & 0.0003\\ & 0.0158\end{bmatrix}$	$\begin{bmatrix}0.0154 & 0.0001\\ & 0.0154\end{bmatrix}$	$\begin{bmatrix}0.0184 & 0.0001\\ & 0.0184\end{bmatrix}$	0.9950	0.9350	0.9947	0.9473	0.9923	0.9690	0.9927	0.9690	0.9920	0.9687
$H_{\mathrm{A}} : \boldsymbol{\gamma} = \begin{pmatrix}0.7\\1.3\end{pmatrix}$	Normal	$\begin{bmatrix}0.7027\\1.3032\end{bmatrix}$	$\begin{bmatrix}0.0227 & 0.0049\\ & 0.0248\end{bmatrix}$	$\begin{bmatrix}0.0219 & 0.0036\\ & 0.0227\end{bmatrix}$	$\begin{bmatrix}0.0147 & 0.0001\\ & 0.0181\end{bmatrix}$	0.8280	0.9393	0.8150	0.9443	0.8573	0.8803	0.8587	0.8730	0.8577	0.8850
	Gamma	$\begin{bmatrix}0.7005\\1.2987\end{bmatrix}$	$\begin{bmatrix}0.0145 & 0.0005\\ & 0.0151\end{bmatrix}$	$\begin{bmatrix}0.0145 & 0.0003\\ & 0.0145\end{bmatrix}$	$\begin{bmatrix}0.0166 & 0.0001\\ & 0.0166\end{bmatrix}$	0.8970	0.9453	0.8860	0.9530	0.8603	0.9647	0.8637	0.9653	0.8573	0.9667
$N = 900$															
$H_0 : \boldsymbol{\gamma} = \begin{pmatrix}1.0\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}1.0033\\1.0003\end{bmatrix}$	$\begin{bmatrix}0.0120 & 0.0026\\ & 0.0112\end{bmatrix}$	$\begin{bmatrix}0.0113 & 0.0024\\ & 0.0109\end{bmatrix}$	$\begin{bmatrix}0.0081 & 0.0000\\ & 0.0080\end{bmatrix}$	0.0603	0.9397	0.0593	0.9407	0.1187	0.8813	0.1203	0.8797	0.1183	0.8817
	Gamma	$\begin{bmatrix}1.0012\\0.9995\end{bmatrix}$	$\begin{bmatrix}0.0074 & 0.0001\\ & 0.0072\end{bmatrix}$	$\begin{bmatrix}0.0073 & 0.0002\\ & 0.0072\end{bmatrix}$	$\begin{bmatrix}0.0083 & 0.0000\\ & 0.0082\end{bmatrix}$	0.0593	0.9407	0.0517	0.9483	0.0357	0.9643	0.0357	0.9643	0.0357	0.9643
$H_{\mathrm{A}} : \boldsymbol{\gamma} = \begin{pmatrix}0.4\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}0.3992\\1.0009\end{bmatrix}$	$\begin{bmatrix}0.0102 & 0.0010\\ & 0.0098\end{bmatrix}$	$\begin{bmatrix}0.0099 & 0.0008\\ & 0.0096\end{bmatrix}$	$\begin{bmatrix}0.0076 & 0.0000\\ & 0.0088\end{bmatrix}$	0.9997	0.9403	0.9993	0.9437	0.9997	0.9120	1.0000	0.9100	0.9997	0.9137
	Gamma	$\begin{bmatrix}0.3984\\0.9997\end{bmatrix}$	$\begin{bmatrix}0.0082 & 0.0002\\ & 0.0076\end{bmatrix}$	$\begin{bmatrix}0.0078 & 0.0001\\ & 0.0077\end{bmatrix}$	$\begin{bmatrix}0.0092 & 0.0000\\ & 0.0091\end{bmatrix}$	1.0000	0.9407	1.0000	0.9477	1.0000	0.9647	1.0000	0.9633	1.0000	0.9657
$H_{\mathrm{A}} : \boldsymbol{\gamma} = \begin{pmatrix}0.7\\1.3\end{pmatrix}$	Normal	$\begin{bmatrix}0.7019\\1.3011\end{bmatrix}$	$\begin{bmatrix}0.0118 & 0.0023\\ & 0.0115\end{bmatrix}$	$\begin{bmatrix}0.0113 & 0.0022\\ & 0.0113\end{bmatrix}$	$\begin{bmatrix}0.0073 & 0.0000\\ & 0.0090\end{bmatrix}$	0.9853	0.9400	0.9847	0.9437	0.9900	0.8793	0.9893	0.8763	0.9890	0.8790
	Gamma	$\begin{bmatrix}0.7008\\1.3005\end{bmatrix}$	$\begin{bmatrix}0.0073 & 0.0003\\ & 0.0070\end{bmatrix}$	$\begin{bmatrix}0.0073 & 0.0001\\ & 0.0072\end{bmatrix}$	$\begin{bmatrix}0.0083 & 0.0000\\ & 0.0082\end{bmatrix}$	0.9953	0.9453	0.9953	0.9510	0.9937	0.9647	0.9937	0.9643	0.9930	0.9963

Table 1. Continued

	Working model	mean($\hat{\boldsymbol{\gamma}}$)	S^2	$\mathrm{Var_A}(\hat{\boldsymbol{\gamma}})$	$\mathrm{Var}(\hat{\boldsymbol{\gamma}})$	AWα	AWcp	ASα	AScp	Lα	Lcp	Wα	Wcp	Sα	Scp
$N = 1350$															
$H_0 : \boldsymbol{\gamma} = \begin{pmatrix}1.0\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}1.0022\\1.0015\end{bmatrix}$	$\begin{bmatrix}0.0076 & 0.0014\\ & 0.0073\end{bmatrix}$	$\begin{bmatrix}0.0073 & 0.0014\\ & 0.0073\end{bmatrix}$	$\begin{bmatrix}0.0054 & 0.0000\\ & 0.0054\end{bmatrix}$	0.0560	0.9440	0.0543	0.9457	0.1123	0.8877	0.1130	0.8870	0.1100	0.8900
	Gamma	$\begin{bmatrix}1.0013\\1.0014\end{bmatrix}$	$\begin{bmatrix}0.0049 & 0.0000\\ & 0.0047\end{bmatrix}$	$\begin{bmatrix}0.0048 & 0.0000\\ & 0.0048\end{bmatrix}$	$\begin{bmatrix}0.0055 & 0.0000\\ & 0.0055\end{bmatrix}$	0.0527	0.9473	0.0450	0.9550	0.0320	0.9680	0.0327	0.9673	0.0323	0.9677
$H_A : \boldsymbol{\gamma} = \begin{pmatrix}0.4\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}0.4003\\1.0033\end{bmatrix}$	$\begin{bmatrix}0.0068 & 0.0004\\ & 0.0068\end{bmatrix}$	$\begin{bmatrix}0.0064 & 0.0005\\ & 0.0065\end{bmatrix}$	$\begin{bmatrix}0.0050 & 0.0000\\ & 0.0059\end{bmatrix}$	1.0000	0.9367	1.0000	0.9393	1.0000	0.9093	1.0000	0.9083	1.0000	0.9093
	Gamma	$\begin{bmatrix}0.4003\\1.0023\end{bmatrix}$	$\begin{bmatrix}0.0054 & -0.0001\\ & 0.0054\end{bmatrix}$	$\begin{bmatrix}0.0052 & 0.0000\\ & 0.0052\end{bmatrix}$	$\begin{bmatrix}0.0061 & 0.0000\\ & 0.0061\end{bmatrix}$	1.0000	0.9410	1.0000	0.9460	1.0000	0.9680	1.0000	0.9670	1.0000	0.9693
$H_A : \boldsymbol{\gamma} = \begin{pmatrix}0.7\\1.3\end{pmatrix}$	Normal	$\begin{bmatrix}0.7002\\1.3016\end{bmatrix}$	$\begin{bmatrix}0.0075 & 0.0013\\ & 0.0078\end{bmatrix}$	$\begin{bmatrix}0.0073 & 0.0012\\ & 0.0076\end{bmatrix}$	$\begin{bmatrix}0.0049 & 0.0000\\ & 0.0060\end{bmatrix}$	0.9983	0.9373	0.9983	0.9430	0.9983	0.8793	0.9983	0.8790	0.9983	0.8813
	Gamma	$\begin{bmatrix}0.6993\\1.3005\end{bmatrix}$	$\begin{bmatrix}0.0049 & 0.0002\\ & 0.0048\end{bmatrix}$	$\begin{bmatrix}0.0048 & 0.0001\\ & 0.0049\end{bmatrix}$	$\begin{bmatrix}0.0055 & 0.0000\\ & 0.0055\end{bmatrix}$	0.9993	0.9427	0.9993	0.9463	0.9993	0.9663	0.9993	0.9623	0.9993	0.9637

Table 2. Model 2: $\mu_i = (2.5\eta_i + 2/3)^3, i = 1, 2, \ldots, N$.

	Working model	mean($\hat{\boldsymbol{\gamma}}$)	S^2	$\mathrm{Var_A}(\hat{\boldsymbol{\gamma}})$	$\mathrm{Var}(\hat{\boldsymbol{\gamma}})$	AWα	AWcp	ASα	AScp	Lα	Lcp	Wα	Wcp	Sα	Scp
$N = 450$															
$H_0 : \boldsymbol{\gamma} = \begin{pmatrix}1.0\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}1.0067\\1.0107\end{bmatrix}$	$\begin{bmatrix}0.0681 & 0.0342\\ & 0.0674\end{bmatrix}$	$\begin{bmatrix}0.0566 & 0.0261\\ & 0.0531\end{bmatrix}$	$\begin{bmatrix}0.0302 & 0.0017\\ & 0.0299\end{bmatrix}$	0.1123	0.8877	0.1090	0.8910	0.1723	0.8277	0.1763	0.8273	0.1650	0.8350
	Gamma	$\begin{bmatrix}0.9903\\0.9946\end{bmatrix}$	$\begin{bmatrix}0.0144 & -0.0001\\ & 0.0141\end{bmatrix}$	$\begin{bmatrix}0.0133 & 0.0002\\ & 0.0132\end{bmatrix}$	$\begin{bmatrix}0.0099 & -0.0003\\ & 0.0099\end{bmatrix}$	0.0877	0.9123	0.0640	0.9360	0.1190	0.8810	0.1200	0.8800	0.1167	0.8833
$H_A : \boldsymbol{\gamma} = \begin{pmatrix}0.4\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}0.3974\\1.0059\end{bmatrix}$	$\begin{bmatrix}0.0221 & 0.0061\\ & 0.0247\end{bmatrix}$	$\begin{bmatrix}0.0209 & 0.0058\\ & 0.0225\end{bmatrix}$	$\begin{bmatrix}0.0132 & 0.0005\\ & 0.0160\end{bmatrix}$	0.9753	0.9027	0.9690	0.9103	0.9767	0.8677	0.9783	0.8650	0.9737	0.8707
	Gamma	$\begin{bmatrix}0.3953\\0.9963\end{bmatrix}$	$\begin{bmatrix}0.0081 & -0.0001\\ & 0.0090\end{bmatrix}$	$\begin{bmatrix}0.0076 & 0.0001\\ & 0.0085\end{bmatrix}$	$\begin{bmatrix}0.0066 & -0.0001\\ & 0.0068\end{bmatrix}$	1.0000	0.9257	1.0000	0.9457	1.0000	0.9093	1.0000	0.9080	1.0000	0.9083
$H_A : \boldsymbol{\gamma} = \begin{pmatrix}0.7\\1.3\end{pmatrix}$	Normal	$\begin{bmatrix}0.7022\\1.3157\end{bmatrix}$	$\begin{bmatrix}0.0637 & 0.0317\\ & 0.0785\end{bmatrix}$	$\begin{bmatrix}0.0553 & 0.0246\\ & 0.0605\end{bmatrix}$	$\begin{bmatrix}0.0283 & 0.0017\\ & 0.0350\end{bmatrix}$	0.6840	0.8790	0.6817	0.8853	0.7183	0.8220	0.7190	0.8197	0.7127	0.8287
	Gamma	$\begin{bmatrix}0.6915\\1.2931\end{bmatrix}$	$\begin{bmatrix}0.0131 & -0.0001\\ & 0.0149\end{bmatrix}$	$\begin{bmatrix}0.0123 & 0.0002\\ & 0.0140\end{bmatrix}$	$\begin{bmatrix}0.0096 & -0.0002\\ & 0.0100\end{bmatrix}$	0.9270	0.9137	0.9107	0.9337	0.9530	0.8760	0.9557	0.8750	0.9497	0.8760
$N = 900$															
$H_0 : \boldsymbol{\gamma} = \begin{pmatrix}1.0\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}1.0094\\1.0093\end{bmatrix}$	$\begin{bmatrix}0.0289 & 0.0139\\ & 0.0271\end{bmatrix}$	$\begin{bmatrix}0.0269 & 0.0131\\ & 0.0268\end{bmatrix}$	$\begin{bmatrix}0.0152 & 0.0009\\ & 0.0152\end{bmatrix}$	0.0910	0.9090	0.0897	0.9103	0.1533	0.8467	0.1560	0.8440	0.1507	0.8493
	Gamma	$\begin{bmatrix}0.9997\\0.9994\end{bmatrix}$	$\begin{bmatrix}0.0068 & 0.0000\\ & 0.0065\end{bmatrix}$	$\begin{bmatrix}0.0068 & 0.0001\\ & 0.0067\end{bmatrix}$	$\begin{bmatrix}0.0049 & -0.0002\\ & 0.0049\end{bmatrix}$	0.0657	0.9343	0.0510	0.9490	0.1047	0.8953	0.1067	0.8933	0.1037	0.8963
$H_A : \boldsymbol{\gamma} = \begin{pmatrix}0.4\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}0.4030\\1.0058\end{bmatrix}$	$\begin{bmatrix}0.0114 & 0.0032\\ & 0.0111\end{bmatrix}$	$\begin{bmatrix}0.0107 & 0.0032\\ & 0.0112\end{bmatrix}$	$\begin{bmatrix}0.0067 & 0.0002\\ & 0.0081\end{bmatrix}$	0.9983	0.9230	0.9967	0.9247	0.9980	0.8730	0.9987	0.8723	0.9980	0.8760
	Gamma	$\begin{bmatrix}0.3998\\0.9998\end{bmatrix}$	$\begin{bmatrix}0.0038 & 0.0000\\ & 0.0041\end{bmatrix}$	$\begin{bmatrix}0.0038 & 0.0032\\ & 0.0043\end{bmatrix}$	$\begin{bmatrix}0.0033 & -0.0001\\ & 0.0034\end{bmatrix}$	1.0000	0.9443	1.0000	0.9500	1.0000	0.9217	1.0000	0.9223	1.0000	0.9223
$H_A : \boldsymbol{\gamma} = \begin{pmatrix}0.7\\1.3\end{pmatrix}$	Normal	$\begin{bmatrix}0.7081\\1.3132\end{bmatrix}$	$\begin{bmatrix}0.0289 & 0.0129\\ & 0.0302\end{bmatrix}$	$\begin{bmatrix}0.0269 & 0.0125\\ & 0.0301\end{bmatrix}$	$\begin{bmatrix}0.0146 & 0.0009\\ & 0.0177\end{bmatrix}$	0.9177	0.9047	0.9190	0.9060	0.9187	0.8457	0.9150	0.8447	0.9183	0.8500
	Gamma	$\begin{bmatrix}0.6998\\1.2995\end{bmatrix}$	$\begin{bmatrix}0.0063 & 0.0000\\ & 0.0071\end{bmatrix}$	$\begin{bmatrix}0.0063 & 0.0001\\ & 0.0072\end{bmatrix}$	$\begin{bmatrix}0.0048 & -0.0002\\ & 0.0049\end{bmatrix}$	0.9983	0.9313	0.9973	0.9497	0.9987	0.8933	0.9990	0.8917	0.9987	0.8910

Table 2. Continued

	Working model	mean($\hat{\boldsymbol{\gamma}}$)	S^2	$\mathrm{Var}_\mathrm{A}(\hat{\boldsymbol{\gamma}})$	$\mathrm{Var}(\hat{\boldsymbol{\gamma}})$	AWα	AWcp	ASα	AScp	Lα	Lcp	Wα	Wcp	Sα	Scp
$N = 1350$															
$H_0 : \boldsymbol{\gamma} = \begin{pmatrix}1.0\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}1.0010\\1.0048\end{bmatrix}$	$\begin{bmatrix}0.0170 & 0.0083\\ & 0.0173\end{bmatrix}$	$\begin{bmatrix}0.0169 & 0.0081\\ & 0.0174\end{bmatrix}$	$\begin{bmatrix}0.0099 & 0.0005\\ & 0.0100\end{bmatrix}$	0.0903	0.9097	0.0897	0.9103	0.1560	0.8440	0.1563	0.8437	0.1513	0.8487
	Gamma	$\begin{bmatrix}0.9982\\1.0005\end{bmatrix}$	$\begin{bmatrix}0.0045 & 0.0002\\ & 0.0044\end{bmatrix}$	$\begin{bmatrix}0.0045 & 0.0001\\ & 0.0045\end{bmatrix}$	$\begin{bmatrix}0.0033 & -0.0001\\ & 0.0033\end{bmatrix}$	0.0610	0.9390	0.0500	0.9500	0.1183	0.8817	0.1167	0.8833	0.1157	0.8843
$H_\mathrm{A} : \boldsymbol{\gamma} = \begin{pmatrix}0.4\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}0.3990\\1.0034\end{bmatrix}$	$\begin{bmatrix}0.0070 & 0.0020\\ & 0.0073\end{bmatrix}$	$\begin{bmatrix}0.0070 & 0.0020\\ & 0.0074\end{bmatrix}$	$\begin{bmatrix}0.0044 & 0.0001\\ & 0.0053\end{bmatrix}$	1.0000	0.9313	1.0000	0.9340	1.0000	0.8737	1.0000	0.8727	1.0000	0.8743
	Gamma	$\begin{bmatrix}0.3994\\1.0006\end{bmatrix}$	$\begin{bmatrix}0.0025 & 0.0001\\ & 0.0028\end{bmatrix}$	$\begin{bmatrix}0.0026 & 0.0001\\ & 0.0029\end{bmatrix}$	$\begin{bmatrix}0.0022 & 0.0000\\ & 0.0023\end{bmatrix}$	1.0000	0.9460	1.0000	0.9530	1.0000	0.9107	1.0000	0.9113	1.0000	0.9103
$H_\mathrm{A} : \boldsymbol{\gamma} = \begin{pmatrix}0.7\\1.3\end{pmatrix}$	Normal	$\begin{bmatrix}0.7001\\1.3077\end{bmatrix}$	$\begin{bmatrix}0.0172 & 0.0078\\ & 0.0193\end{bmatrix}$	$\begin{bmatrix}0.0172 & 0.0077\\ & 0.0193\end{bmatrix}$	$\begin{bmatrix}0.0095 & 0.0005\\ & 0.0116\end{bmatrix}$	0.9850	0.9123	0.9870	0.9160	0.9880	0.8430	0.9863	0.8400	0.9877	0.8450
	Gamma	$\begin{bmatrix}0.6986\\1.3008\end{bmatrix}$	$\begin{bmatrix}0.0042 & 0.0002\\ & 0.0048\end{bmatrix}$	$\begin{bmatrix}0.0042 & 0.0001\\ & 0.0049\end{bmatrix}$	$\begin{bmatrix}0.0032 & -0.0001\\ & 0.0033\end{bmatrix}$	0.9997	0.9417	0.9997	0.9497	1.0000	0.8907	1.0000	0.8917	1.0000	0.8897

Table 3. Model 3: $\mu_i = \eta_i^2, i = 1, 2, \dots, N$.

	Working model	mean($\hat{\boldsymbol{\gamma}}$)	S^2	$\mathrm{Var}_{\mathrm{A}}(\hat{\boldsymbol{\gamma}})$	$\mathrm{Var}(\hat{\boldsymbol{\gamma}})$	AWα	AWcp	ASα	AScp	Lα	Lcp	Wα	Wcp	Sα	Scp
$N = 450$															
$H_0 : \boldsymbol{\gamma} = \begin{pmatrix}1.0\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}1.0023\\0.9991\end{bmatrix}$	$\begin{bmatrix}0.0242 & 0.0040\\ & 0.0232\end{bmatrix}$	$\begin{bmatrix}0.0237 & 0.0038\\ & 0.0236\end{bmatrix}$	$\begin{bmatrix}0.0203 & 0.0009\\ & 0.0203\end{bmatrix}$	0.0573	0.9427	0.0527	0.9473	0.0793	0.9207	0.0810	0.9190	0.0777	0.9223
	Gamma	$\begin{bmatrix}1.0020\\0.9989\end{bmatrix}$	$\begin{bmatrix}0.0162 & -0.0001\\ & 0.0163\end{bmatrix}$	$\begin{bmatrix}0.0163 & -0.0003\\ & 0.0164\end{bmatrix}$	$\begin{bmatrix}0.0191 & -0.0008\\ & 0.0192\end{bmatrix}$	0.0487	0.9513	0.0377	0.9623	0.0270	0.9730	0.0263	0.9737	0.0293	0.9707
$H_{\mathrm{A}} : \boldsymbol{\gamma} = \begin{pmatrix}0.4\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}0.3966\\0.9971\end{bmatrix}$	$\begin{bmatrix}0.0180 & 0.0015\\ & 0.0183\end{bmatrix}$	$\begin{bmatrix}0.0181 & 0.0014\\ & 0.0178\end{bmatrix}$	$\begin{bmatrix}0.0154 & 0.0004\\ & 0.0166\end{bmatrix}$	0.9833	0.9460	0.9813	0.9490	0.9877	0.9310	0.9880	0.9307	0.9873	0.9343
	Gamma	$\begin{bmatrix}0.3977\\0.9968\end{bmatrix}$	$\begin{bmatrix}0.0129 & -0.0004\\ & 0.0137\end{bmatrix}$	$\begin{bmatrix}0.0133 & -0.0002\\ & 0.0134\end{bmatrix}$	$\begin{bmatrix}0.0155 & -0.0003\\ & 0.0162\end{bmatrix}$	0.9980	0.9370	0.9960	0.9550	0.9967	0.9697	0.9967	0.9690	0.9967	0.9680
$H_{\mathrm{A}} : \boldsymbol{\gamma} = \begin{pmatrix}0.7\\1.3\end{pmatrix}$	Normal	$\begin{bmatrix}0.7001\\1.3000\end{bmatrix}$	$\begin{bmatrix}0.0244 & 0.0041\\ & 0.0256\end{bmatrix}$	$\begin{bmatrix}0.0243 & 0.0034\\ & 0.0242\end{bmatrix}$	$\begin{bmatrix}0.0199 & 0.0008\\ & 0.0215\end{bmatrix}$	0.7823	0.9383	0.7750	0.9413	0.7927	0.9187	0.7927	0.9163	0.7887	0.9217
	Gamma	$\begin{bmatrix}0.7001\\1.2998\end{bmatrix}$	$\begin{bmatrix}0.0158 & -0.0004\\ & 0.0170\end{bmatrix}$	$\begin{bmatrix}0.0161 & -0.0003\\ & 0.0164\end{bmatrix}$	$\begin{bmatrix}0.0187 & -0.0007\\ & 0.0195\end{bmatrix}$	0.8360	0.9413	0.8287	0.9593	0.7793	0.9757	0.7850	0.9740	0.7723	0.9730
$N = 900$															
$H_0 : \boldsymbol{\gamma} = \begin{pmatrix}1.0\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}1.0004\\1.0010\end{bmatrix}$	$\begin{bmatrix}0.0128 & 0.0019\\ & 0.0117\end{bmatrix}$	$\begin{bmatrix}0.0121 & 0.0020\\ & 0.0117\end{bmatrix}$	$\begin{bmatrix}0.0102 & 0.0004\\ & 0.0101\end{bmatrix}$	0.0567	0.9433	0.0553	0.9447	0.0743	0.9257	0.0757	0.9243	0.0723	0.9277
	Gamma	$\begin{bmatrix}1.0003\\1.0023\end{bmatrix}$	$\begin{bmatrix}0.0087 & -0.0003\\ & 0.0082\end{bmatrix}$	$\begin{bmatrix}0.0083 & -0.0002\\ & 0.0082\end{bmatrix}$	$\begin{bmatrix}0.0096 & -0.0005\\ & 0.0095\end{bmatrix}$	0.0627	0.9373	0.0503	0.9497	0.0367	0.9633	0.0380	0.9620	0.0350	0.9650
$H_{\mathrm{A}} : \boldsymbol{\gamma} = \begin{pmatrix}0.4\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}0.4011\\1.0015\end{bmatrix}$	$\begin{bmatrix}0.0095 & 0.0008\\ & 0.0089\end{bmatrix}$	$\begin{bmatrix}0.0092 & 0.0008\\ & 0.0088\end{bmatrix}$	$\begin{bmatrix}0.0077 & 0.0002\\ & 0.0082\end{bmatrix}$	1.0000	0.9477	1.0000	0.9480	1.0000	0.9313	1.0000	0.9317	1.0000	0.9337
	Gamma	$\begin{bmatrix}0.4029\\1.0022\end{bmatrix}$	$\begin{bmatrix}0.0069 & -0.0002\\ & 0.0069\end{bmatrix}$	$\begin{bmatrix}0.0067 & -0.0001\\ & 0.0067\end{bmatrix}$	$\begin{bmatrix}0.0077 & -0.0002\\ & 0.0080\end{bmatrix}$	1.0000	0.9383	1.0000	0.9513	1.0000	0.9717	1.0000	0.9703	1.0000	0.9697
$H_{\mathrm{A}} : \boldsymbol{\gamma} = \begin{pmatrix}0.7\\1.3\end{pmatrix}$	Normal	$\begin{bmatrix}0.7019\\1.3034\end{bmatrix}$	$\begin{bmatrix}0.0130 & 0.0017\\ & 0.0115\end{bmatrix}$	$\begin{bmatrix}0.0125 & 0.0019\\ & 0.0120\end{bmatrix}$	$\begin{bmatrix}0.0100 & 0.0004\\ & 0.0107\end{bmatrix}$	0.9743	0.9450	0.9747	0.9487	0.9783	0.9207	0.9780	0.9227	0.9780	0.9247
	Gamma	$\begin{bmatrix}0.7015\\1.3033\end{bmatrix}$	$\begin{bmatrix}0.0081 & 0.0000\\ & 0.0082\end{bmatrix}$	$\begin{bmatrix}0.0082 & -0.0002\\ & 0.0082\end{bmatrix}$	$\begin{bmatrix}0.0093 & -0.0005\\ & 0.0096\end{bmatrix}$	0.9897	0.9433	0.9910	0.9503	0.9863	0.9673	0.9870	0.9647	0.9853	0.9683

Table 3. Continued

	Working model	mean($\hat{\boldsymbol{\gamma}}$)	S^2	$\mathrm{Var_A}(\hat{\boldsymbol{\gamma}})$	$\mathrm{Var}(\hat{\boldsymbol{\gamma}})$	AWα	AWcp	ASα	AScp	Lα	Lcp	Wα	Wcp	Sα	Scp
$N = 1350$															
$H_0 : \boldsymbol{\gamma} = \begin{pmatrix}1.0\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}1.0025\\1.0020\end{bmatrix}$	$\begin{bmatrix}0.0082 & 0.0013\\ & 0.0080\end{bmatrix}$	$\begin{bmatrix}0.0080 & 0.0013\\ & 0.0080\end{bmatrix}$	$\begin{bmatrix}0.0068 & 0.0002\\ & 0.0068\end{bmatrix}$	0.0627	0.9373	0.0603	0.9397	0.0860	0.9140	0.0860	0.9140	0.0857	0.9143
	Gamma	$\begin{bmatrix}1.0007\\0.9999\end{bmatrix}$	$\begin{bmatrix}0.0058 & -0.0001\\ & 0.0054\end{bmatrix}$	$\begin{bmatrix}0.0055 & -0.0001\\ & 0.0055\end{bmatrix}$	$\begin{bmatrix}0.0064 & -0.0003\\ & 0.0064\end{bmatrix}$	0.0580	0.9420	0.0507	0.9493	0.0307	0.9693	0.0340	0.9660	0.0303	0.9697
$H_{\mathrm{A}} : \boldsymbol{\gamma} = \begin{pmatrix}0.4\\1.0\end{pmatrix}$	Normal	$\begin{bmatrix}0.4021\\1.0004\end{bmatrix}$	$\begin{bmatrix}0.0061 & 0.0005\\ & 0.0063\end{bmatrix}$	$\begin{bmatrix}0.0061 & 0.0005\\ & 0.0060\end{bmatrix}$	$\begin{bmatrix}0.0051 & 0.0001\\ & 0.0055\end{bmatrix}$	1.0000	0.9423	1.0000	0.9427	1.0000	0.9243	1.0000	0.9243	1.0000	0.9253
	Gamma	$\begin{bmatrix}0.4028\\1.0003\end{bmatrix}$	$\begin{bmatrix}0.0046 & 0.0000\\ & 0.0047\end{bmatrix}$	$\begin{bmatrix}0.0045 & -0.0001\\ & 0.0045\end{bmatrix}$	$\begin{bmatrix}0.0051 & -0.0001\\ & 0.0054\end{bmatrix}$	1.0000	0.9380	1.0000	0.9447	1.0000	0.9660	1.0000	0.9650	1.0000	0.9643
$H_{\mathrm{A}} : \boldsymbol{\gamma} = \begin{pmatrix}0.7\\1.3\end{pmatrix}$	Normal	$\begin{bmatrix}0.7009\\1.3015\end{bmatrix}$	$\begin{bmatrix}0.0085 & 0.0009\\ & 0.0081\end{bmatrix}$	$\begin{bmatrix}0.0082 & 0.0012\\ & 0.0082\end{bmatrix}$	$\begin{bmatrix}0.0066 & 0.0002\\ & 0.0072\end{bmatrix}$	0.9960	0.9453	0.9957	0.9473	0.9950	0.9157	0.9950	0.9150	0.9947	0.9187
	Gamma	$\begin{bmatrix}0.7007\\1.3014\end{bmatrix}$	$\begin{bmatrix}0.0056 & -0.0002\\ & 0.0055\end{bmatrix}$	$\begin{bmatrix}0.0055 & -0.0001\\ & 0.0055\end{bmatrix}$	$\begin{bmatrix}0.0062 & -0.0003\\ & 0.0064\end{bmatrix}$	0.9990	0.9480	0.9993	0.9513	0.9983	0.9707	0.9987	0.9687	0.9987	0.9710

by $\mathrm{Var}(\hat{\boldsymbol{\gamma}})$ and the average of the adjusted covariance matrix estimate of $\hat{\boldsymbol{\gamma}}$, namely $\hat{\boldsymbol{A}}^{-1}\hat{\boldsymbol{B}}\hat{\boldsymbol{A}}^{-1}/N$, denoted by $\mathrm{Var}_{\mathrm{A}}(\hat{\boldsymbol{\gamma}})$ are also included. Note that because with large samples the adjusted Wald and score statistics under the normal and gamma working models are free of σ^2 and r, the non-regression parameters σ^2 and r in the RNR and RGR models are treated known, *a priori*, as the same arbitrarily chosen value of 1, respectively.

From Tables 1–3, it is evident that the adjusting matrices successfully correct the normal and gamma working models and make them robust. As can be seen from Tables 1–3, the averages of the adjusted covariance matrix estimates, $\mathrm{Var}_{\mathrm{A}}(\hat{\boldsymbol{\gamma}})$, are nearly equivalent to the sample covariance matrix of $\hat{\boldsymbol{\gamma}}$, S^2, whereas the averages of the unadjusted covariance matrix estimates, $\mathrm{Var}(\hat{\boldsymbol{\gamma}})$, are different from S^2.

It is also observed that when the simulated data sets are generated under the null hypothesis H_0, the adjusted Wald and score statistics are more effective than the test statistics $Q_{\mathrm{L}}, Q_{\mathrm{W}}$ and Q_{S} in providing the correct type I error probabilities. As can be seen from Tables 1–3, when the data are generated from H_0, the values of AWα and ASα are more close to the nominal significance level 0.05, in contrast with the values of Lα, Wα and Sα.

On the other hand, it is noted that when the simulated data sets are generated under the alternative hypothesis H_{A}, the adjusted Wald and score statistics not only rightly reject the null hypothesis H_0 but also provide the right confidence region. As can be seen from Tables 1–3, when the data are generated from H_{A}, the values of AWα and ASα gradually approach the value of 1 and the values of AWcp and AScp inchmeal approximate to the nominal confidence level 0.95, as the sample size N increases. On the contrary, the test statistics $Q_{\mathrm{L}}, Q_{\mathrm{W}}$ and Q_{S}, under H_{A}, only succeed in rejecting H_0, but they do not provide the exactly correct confidence region. For example, in the case of Model 2 with the sample size $N = 1350$ and the alternative hypothesis $H_{\mathrm{A}} : \boldsymbol{\gamma}^t = (0.7, 1.3)$, respectively, the values of Lcp, Wcp and Scp under the gamma working model, 0.8907, 0.8917, and 0.8897, are far from the nominal confidence level 0.95, in comparison with the values of AWcp and AScp under the gamma working model, 0.9417 and 0.9497.

Obviously, from the results of Tables 1–3, it is enough to verify that the adjusted Wald and score statistics based on the normal and gamma working models furnish a foundation for valid inferences for the regression parameters of interest, even though the true underlying distributions are not from these two working models. Despite the fact that the RNR and RGR models remain the robustness property in misspecified models, some finite sample differences are revealed in the numerical performances.

The results in Tables 1–3 apparently display that the adjusted covariance matrix estimates, $\mathrm{Var}(\hat{\boldsymbol{\gamma}})$, under the gamma working model are smaller than that under the normal working model. This explicitly means that in terms of the hypothesis testing, the RGR model is more powerful than the RNR model for non-negative continuous data.

5. Concluding remarks

We propose the parametric robust regression methods in the GLM setting. The proposed methods can provide the valid inferences about the regression parameters of interest under model misspecification.

The adjusting matrices for the normal and gamma working models are submitted here. They successfully adjust these two working models into robust models, whatever the true underlying distributions are, as long as their second moments exist. The two adjusted models, namely the RNR and RGR models, warrant the asymptotically legitimate inferences under model misspecification. Simulation studies illustrate that the RGR model is more efficient for more general non-negative continuous random variables.

One of the many attractive features of our proposed methods is that with large samples, the effect of σ^2 of the normal model and the effect of r of the gamma model are entirely purged by their respective adjustments. Hence, although the non-regression parameters σ^2 and r are artificially given positive values, the asymptotic validity of the RNR and RGR models are always obtained.

Finally, we noted that the above discussion was centred on the case of all continuous random variables. For robust inferences for count data, we refer the interested readers to [21].

Acknowledgements

The authors are grateful for helpful comments from the associate editor and the reviewer, which improved the quality of this work.

References

[1] J.A. Nelder and R.W.M. Wedderburn, *Generalized linear models*, J. R. Statist. Soc. Ser. A 135 (1972), pp. 370–384.
[2] L. Fahrmeir and H. Kaufmann, *Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models*, Ann. Stat. 13 (1985), pp. 342–368.
[3] L. Fahrmeir, *Asymptotic testing theory for generalized linear models*, Statistics 18 (1987), pp. 65–76.
[4] L. Fahrmeir, *Maximum likelihood estimation in misspecified generalized linear models*, Statistics 21 (1990), pp. 487–502.
[5] E. Cantoni and E. Ronchetti, *Robust inference for generalized linear models*, J. Am. Statist. Assoc. 96 (2001), pp. 1022–1030.
[6] C. Adimari and L. Ventura, *Robust inference for generalized linear models with application to logistic regression*, Statist. Probab. Lett. 55 (2001), pp. 413–419.
[7] T. Li and C. Hsiao, *Robust estimation of generalized linear models with measurement errors*, J. Econom. 118 (2004), pp. 51–65.
[8] S.K. Sinha, *Robust methods for generalized linear models with nonignorable missing covariates*, Canad. J. Stat. 36 (2008), pp. 277–299.
[9] A.M. Bianco, G. Boente, and I.M. Rodrigues, *Robust tests in generalized linear models with missing responses*, Comput. Stat. Data Anal. (in press), doi: 10.1016.csda.2012.05.008.
[10] P.J. Heagerty and B.F. Kurland, *Misspecified maximum likelihood estimates and generalised linear mixed models*, Biometrika 88 (2001), pp. 973–985.
[11] J. Jiang and W. Zhang, *Robust estimation in generalised linear mixed models*, Biometrika 88 (2001), pp. 753–765.
[12] K.K.W. Yau, and A.Y.C. Kuk, *Robust estimation in generalized linear mixed models*, J. R. Statist. Soc. Ser. B 64 (2002), pp. 101–117.
[13] S.K. Sinha, *Robust analysis of generalized linear mixed models*, J. Am. Statist. Assoc. 99 (2004), pp. 451–460.
[14] S.K. Sinha, *Robust inference in generalized linear models for longitudinal data*, Canad. J. Stat. 34 (2006), pp. 261–278.
[15] A.M. Richardson and A.H. Welsh, *Robust restricted maximum likelihood in mixed linear models*, Biometrics 51 (1995), pp. 1429–1439.
[16] S. Yun and Y. Lee, *Robust estimation in mixed linear models with non-monotone missingness*, Stat. Med. 25 (2006), pp. 3877–3892.
[17] H. Jacqmin-Gadda, S. Sibillot, C. Proust, J.-M. Molina, and R. Thiebaut, *Robustness of the linear mixed model to misspecified error distribution*, Comput. Stat. Data Anal. 51 (2007), pp. 5142–5154.
[18] R.M. Royall and T.-S. Tsou, *Interpreting statistical evidence using imperfect models: Robust adjusted likelihood functions*, J. R. Statist. Soc. Ser. B 65 (2003), pp. 391–404.
[19] T.-S. Tsou, *Comparing two population means and variances – a parametric robust way*, Comm. Stat. Theory Methods 32 (2003), pp. 2013–2019.
[20] T.-S. Tsou and K.-F. Cheng, *Parametric robust regression analysis of contaminated data*, Comm. Stat. Theory Methods 33 (2004), pp. 1887–1898.
[21] T.-S. Tsou, *Robust Poisson regression*, J. Statist. Plann. Inference 136 (2006), pp. 3173–3186.
[22] H. White, *Maximum likelihood estimation of misspecified models*, Econometrica 50 (1982), pp. 1–25.
[23] R.W.M. Wedderburn, *Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method*, Biometrika 61 (1974), pp. 439–447.
[24] P. McCullagh and J.A. Nelder, *Generalized Linear Models*, Chapman and Hall, New York, 1989.
[25] C.C. Heyde, *Quasi-likelihood and Its Application*, Springer-Verlag, New York, 1997.
[26] P.J. Huber, *Robust Statistics*, John Wiley, New York, 1981.
[27] P. McCullagh, *Quasi-likelihood functions*, Ann. Stat. 11 (1983), pp. 59–67.
[28] D.R. Cox and D.V. Hinkley, *Theoretical Statistics*, Chapman and Hall, New York, 1996.

Appendix

Here some details regarding the quantities required to calculate the adjusting matrices $\boldsymbol{A}_n$ and $\boldsymbol{B}_n$ of $\boldsymbol{A}_n\boldsymbol{B}_n^{-1}\boldsymbol{A}_n$ that correct the normal regression model are provided.

To facilitate calculation of the adjusting matrices, let $\boldsymbol{\beta}^t = (\beta_1, \beta_2, \ldots, \beta_w) = (\gamma_{p-1}, \gamma_{p-2}, \ldots, \gamma_{p-w})$ be the w-vector of parameters of interest and let $\boldsymbol{\varphi}$ be the $(p-w+1)$-dimensional nuisance parameters with the non-regression parameter σ^2 and the $(p-w)$ regression coefficients $(\gamma_0, \gamma_1, \ldots, \gamma_{p-w-1})$. Under the normal working model, $\boldsymbol{I}_{h\boldsymbol{\beta\beta}}$ and $\boldsymbol{I}_{h\boldsymbol{\beta\varphi}}$ are approximately equal to the $w \times w$ matrix

$$\frac{1}{N\sigma_0^2}\begin{bmatrix} \sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,p-1}^2 & \cdots & \sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,p-1}x_{i,p-w} \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,p-w}x_{i,p-1} & \cdots & \sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,p-w}^2 \end{bmatrix}$$

and the $w \times (p-w+1)$ matrix

$$\frac{1}{N\sigma_0^2}\begin{bmatrix} 0 & \sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,p-1}x_{i,0} & \cdots & \sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,p-1}x_{i,p-w-1} \\ 0 & \sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,p-2}x_{i,0} & \cdots & \sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,p-2}x_{i,p-w-1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,p-w}x_{i,0} & \cdots & \sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,p-w}x_{i,p-w-1} \end{bmatrix},$$

respectively. $\boldsymbol{I}_{h\boldsymbol{\varphi\varphi}}$ is asymptotically expressed by

$$\frac{1}{N\sigma_0^2}\begin{bmatrix} i_{\boldsymbol{\varphi\varphi}1} & 0 & \cdots & 0 \\ 0 & \sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,0}^2 & \cdots & \sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,0}x_{i,p-w-1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,p-w-1}x_{i,0} & \cdots & \sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,p-w-1}^2 \end{bmatrix},$$

where $i_{\boldsymbol{\varphi\varphi}1} = -\sigma_0^2 l_{\sigma^2\sigma^2}$.

For simplicity of notation, let $\boldsymbol{\Delta}_n$ be the $(p-w)\times(p-w)$ matrix with the jth row as $(\sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,j-1}x_{i,0}, \ldots, \sum_{i=1}^{N}(\mu'_{i,0})^2 x_{i,j-1}x_{i,p-w-1})$ for $j = 1, 2, \ldots, p-w$. Then, $\boldsymbol{I}_{h\boldsymbol{\varphi\varphi}}$ is approximately written in the form

$$\boldsymbol{I}_{h\boldsymbol{\varphi\varphi}} \approx \frac{1}{N\sigma_0^2}\begin{bmatrix} i_{\boldsymbol{\varphi\varphi}1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_n \end{bmatrix},$$

where $\mathbf{0}$ is a $(p-w)$-vector and consists of only zeros. Its inverse $\boldsymbol{I}_{h\boldsymbol{\varphi\varphi}}^{-1}$ is approximately given by

$$\boldsymbol{I}_{h\boldsymbol{\varphi\varphi}}^{-1} \approx \frac{N\sigma_0^2}{|\boldsymbol{\Delta}_n|}\begin{bmatrix} \frac{|\boldsymbol{\Delta}_n|}{i_{\boldsymbol{\varphi\varphi}1}} & 0 & \cdots & 0 \\ 0 & R_{1,1} & \cdots & R_{p-w,1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & R_{1,p-w} & \cdots & R_{p-w,p-w} \end{bmatrix},$$

where $R_{l,m} = (-1)^{l+m}|M_{(lm)}(\boldsymbol{\Delta}_n)|$ is the (l,m)th cofactor of $\boldsymbol{\Delta}_n$.

Similarly, $\boldsymbol{V}_{h\boldsymbol{\beta\beta}}$, $\boldsymbol{V}_{h\boldsymbol{\beta\varphi}}$ and $\boldsymbol{V}_{h\boldsymbol{\varphi\varphi}}$ are approximately equal to

$$\frac{1}{N\sigma_0^4}\begin{bmatrix} \sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu'_{i,0})^2x_{i,p-1}^2 & \dots & \sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu'_{i,0})^2x_{i,p-1}x_{i,p-w} \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu'_{i,0})^2x_{i,p-w}x_{i,p-1} & \dots & \sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu'_{i,0})^2x_{i,p-w}^2 \end{bmatrix},$$

$$\frac{1}{N\sigma_0^4}\begin{bmatrix} 0 & \sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu'_{i,0})^2x_{i,p-1}x_{i,0} & \dots & \sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu'_{i,0})^2x_{i,p-1}x_{i,p-w-1} \\ 0 & \sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu'_{i,0})^2x_{i,p-2}x_{i,0} & \dots & \sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu'_{i,0})^2x_{i,p-2}x_{i,p-w-1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu'_{i,0})^2x_{i,p-w}x_{i,0} & \dots & \sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu'_{i,0})^2x_{i,p-w}x_{i,p-w-1} \end{bmatrix}$$

and

$$\frac{1}{N\sigma_0^4}\begin{bmatrix} v_{\boldsymbol{\varphi\varphi}1} & v_{\sigma^2\gamma_0} & \dots & v_{\sigma^2\gamma_{p-w-1}} \\ v_{\gamma_0\sigma^2} & \sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu'_{i,0})^2x_{i,0}^2 & \dots & \sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu'_{i,0})^2x_{i,0}x_{i,p-w-1} \\ \vdots & \vdots & \vdots & \vdots \\ v_{\gamma_{p-w-1}\sigma^2} & \sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu'_{i,0})^2x_{i,p-w-1}x_{i,0} & \dots & \sum_{i=1}^{N}\mathrm{Var}_h(Y_i)(\mu'_{i,0})^2x_{i,p-w-1}^2 \end{bmatrix},$$

respectively. Here $v_{\boldsymbol{\varphi\varphi}1} = \sigma_0^4 E_h[l_{\sigma^2}(\boldsymbol{\beta}_0,\boldsymbol{\varphi}_0)l_{\sigma^2}(\boldsymbol{\beta}_0,\boldsymbol{\varphi}_0)]$ and $v_{\sigma^2\gamma_j} = v_{\gamma_j\sigma^2} = \sigma_0^4 E_h[l_{\sigma^2}(\boldsymbol{\beta}_0,\boldsymbol{\varphi}_0)l_{\gamma_j}(\boldsymbol{\beta}_0,\boldsymbol{\varphi}_0)], j = 0, 1, \dots, p-w-1$.

According to Equation (1), the (u, v) entries of $\boldsymbol{A}_n$ are derived as follows:

$$\begin{aligned} \boldsymbol{A}_{n(uv)} &= I_{h\beta_u\beta_v} - \boldsymbol{I}_{h\beta_u\boldsymbol{\varphi}}\boldsymbol{I}^{-1}_{h\boldsymbol{\varphi\varphi}}\boldsymbol{I}_{h\boldsymbol{\varphi}\beta_v} - \boldsymbol{I}_{h\beta_v\boldsymbol{\varphi}}\boldsymbol{I}^{-1}_{h\boldsymbol{\varphi\varphi}}\boldsymbol{I}_{h\boldsymbol{\varphi}\beta_u} + \boldsymbol{I}_{h\beta_v\boldsymbol{\varphi}}\boldsymbol{I}^{-1}_{h\boldsymbol{\varphi\varphi}}\boldsymbol{I}_{h\boldsymbol{\varphi\varphi}}\boldsymbol{I}^{-1}_{h\boldsymbol{\varphi\varphi}}\boldsymbol{I}_{h\boldsymbol{\varphi}\beta_u} \\ &= \lim_{N\to\infty}\frac{1}{N\sigma_0^2}\left\{\sum_{i=1}^{N}(\mu'_{i,0})^2x_{i,p-u}x_{i,p-v} - \frac{1}{|\boldsymbol{\Delta}_n|}\sum_{i=1}^{N}(\mu'_{i,0})^2\left(x_{i,p-v}\sum_{j=1}^{p-w}x_{i,j-1}|\boldsymbol{\Delta}_{nj(u)}|\right)\right. \\ &\quad - \frac{1}{|\boldsymbol{\Delta}_n|}\sum_{i=1}^{N}(\mu'_{i,0})^2\left(x_{i,p-u}\sum_{j=1}^{p-w}x_{i,j-1}|\boldsymbol{\Delta}_{nj(v)}|\right) \\ &\quad \left. + \frac{1}{|\boldsymbol{\Delta}_n|^2}\sum_{i=1}^{N}(\mu'_{i,0})^2\left(\sum_{j=1}^{p-w}x_{i,j-1}|\boldsymbol{\Delta}_{nj(v)}|\right)\left(\sum_{j=1}^{p-w}x_{i,j-1}|\boldsymbol{\Delta}_{nj(u)}|\right)\right\} \\ &= \lim_{N\to\infty}\frac{1}{N\sigma_0^2}\sum_{i=1}^{N}(\mu'_{i,0})^2\left(x_{i,p-u} - \sum_{j=1}^{p-w}\frac{|\boldsymbol{\Delta}_{nj(u)}|}{|\boldsymbol{\Delta}_n|}x_{i,j-1}\right)\left(x_{i,p-v} - \sum_{j=1}^{p-w}\frac{|\boldsymbol{\Delta}_{nj(v)}|}{|\boldsymbol{\Delta}_n|}x_{i,j-1}\right), \end{aligned}$$

where $|\boldsymbol{\Delta}_{nj(u)}| = \sum_{m=1}^{p-w}(R_{j,m}\sum_{i=1}^{N}(\mu'_{i,0})^2x_{i,m-1}x_{i,p-u})$.

According to Equation (2), the (u, v) entries of $\boldsymbol{B}_n$ are derived as follows:

$$\begin{aligned}
\boldsymbol{B}_{n(uv)} &= V_{h\beta_u\beta_v} - \boldsymbol{I}_{h\beta_u\boldsymbol{\varphi}}\boldsymbol{I}^{-1}_{h\boldsymbol{\varphi}\boldsymbol{\varphi}}\boldsymbol{V}_{h\boldsymbol{\varphi}\beta_v} - \boldsymbol{V}_{h\beta_u\boldsymbol{\varphi}}\boldsymbol{I}^{-1}_{h\boldsymbol{\varphi}\boldsymbol{\varphi}}\boldsymbol{I}_{h\boldsymbol{\varphi}\beta_v} + \boldsymbol{I}_{h\beta_u\boldsymbol{\varphi}}\boldsymbol{I}^{-1}_{h\boldsymbol{\varphi}\boldsymbol{\varphi}}\boldsymbol{V}_{h\boldsymbol{\varphi}\boldsymbol{\varphi}}\boldsymbol{I}^{-1}_{h\boldsymbol{\varphi}\boldsymbol{\varphi}}\boldsymbol{I}_{h\boldsymbol{\varphi}\beta_v} \\
&= \lim_{N\to\infty} \frac{1}{N\sigma_0^4} \left\{ \sum_{i=1}^{N} \mathrm{Var}_h(Y_i)(\mu'_{i,0})^2 x_{i,p-u}x_{i,p-v} - \frac{1}{|\boldsymbol{\Delta}_n|}\sum_{i=1}^{N} \mathrm{Var}_h(Y_i)(\mu'_{i,0})^2 \left(x_{i,p-v}\sum_{j=1}^{p-w} x_{i,j-1}|\boldsymbol{\Delta}_{nj(u)}| \right) \right. \\
&\quad - \frac{1}{|\boldsymbol{\Delta}_n|}\sum_{i=1}^{N} \mathrm{Var}_h(Y_i)(\mu'_{i,0})^2 \left(x_{i,p-u}\sum_{j=1}^{p-w} x_{i,j-1}|\boldsymbol{\Delta}_{nj(v)}| \right) \\
&\quad \left. + \frac{1}{|\boldsymbol{\Delta}_n|^2}\sum_{i=1}^{N} \mathrm{Var}_h(Y_i)(\mu'_{i,0})^2 \left(\sum_{j=1}^{p-w} x_{i,j-1}|\boldsymbol{\Delta}_{nj(u)}| \right)\left(\sum_{j=1}^{p-w} x_{i,j-1}|\boldsymbol{\Delta}_{nj(v)}| \right) \right\} \\
&= \lim_{N\to\infty} \frac{1}{N\sigma_0^4} \sum_{i=1}^{N} \mathrm{Var}_h(Y_i)(\mu'_{i,0})^2 \left(x_{i,p-u} - \sum_{j=1}^{p-w} \frac{|\boldsymbol{\Delta}_{nj(u)}|}{|\boldsymbol{\Delta}_n|} x_{i,j-1} \right)\left(x_{i,p-v} - \sum_{j=1}^{p-w} \frac{|\boldsymbol{\Delta}_{nj(v)}|}{|\boldsymbol{\Delta}_n|} x_{i,j-1} \right).
\end{aligned}$$

The adjusting matrices $\boldsymbol{A}_g$ and $\boldsymbol{B}_g$ of the gamma working model are derived in the analogous way.