Parametric simultaneous robust inferences for regression coefficient under generalized linear models

Li-Chu Chien\textsuperscript{a} & Tsung-Shan Tsou\textsuperscript{bcd}

\textsuperscript{a} Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan, Republic of China
\textsuperscript{b} Institute of Statistics, National Central University, Jhongli, Taiwan, Republic of China
\textsuperscript{c} Institute of Systems Biology and Bioinformatics, Center for Biotechnology and Biomedical Engineering, National Central University, Jhongli, Taiwan, Republic of China
\textsuperscript{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
Parametric simultaneous robust inferences for regression coefficient under generalized linear models

Li-Chu Chiena* and Tsung-Shan Tsoub,c,d

aInstitute of Statistics, National Chiao Tung University, Hsinchu, Taiwan, Republic of China; bInstitute of Statistics, National Central University, Jhongli, Taiwan, Republic of China; cInstitute of Systems Biology and Bioinformatics, Center for Biotechnology and Biomedical Engineering, National Central University, Jhongli, Taiwan, Republic of China; dCathay 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. 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
© 2012 Taylor & Francis
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.


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 \( Y_1, Y_2, \ldots, Y_N \) 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, \( f_i = f_i(\cdot; \psi) = f(\cdot; \eta_i(\psi)) \), \( i = 1, 2, \ldots, N, \psi \in \Psi \), where \( \psi \) is a fixed-dimensional vector of unknown parameters. For example, under normal regression settings, \( \eta_i(\psi) = (x_i' \gamma, \sigma^2) \), \( \psi = (\psi', \sigma^2)' \) and \( f_i = f_i(y_i; \psi) = \exp(-(y_i - x_i' \gamma)^2 / 2\sigma^2) / \sqrt{2\pi} \sigma \). Here \( x_i \) represents the \( p \) characteristics that are specific to \( y_i \), and \( \gamma \) represents the \( p \) regression coefficients that describe how \( x_i \) 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 \( \psi \).

Now partition \( \psi \) as \( \psi' = (\theta', \varphi') \), where \( \theta \) is the \( p \)-vector of parameters of interest and \( \varphi \) is the remaining fixed-dimensional nuisance parameters. Let \( \theta_0 \) and \( \varphi_0 \) denote the limiting values of
the MLEs, $\hat{\Theta}$ and $\hat{\Phi}$, based on the working model $f = (f_1, f_2, \ldots, f_N)$, when the $Y_i$’s are actually generated from the family $\{h_i = h(\bullet, \tau_i(\Theta, \Lambda)), i = 1, 2, \ldots, N\}$, where $\Lambda$ 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 $\Theta$, remain the parameters of interest under $h$, so that $\Theta_0$ 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 $l_\Theta$ and $l_\Phi$ for the first derivatives of the log-likelihood function $l(\Theta, \Phi)$ w.r.t. $\Theta$ and $\Phi$, respectively, whose derivatives w.r.t. $\Theta$ are correspondingly denoted by $l_{\Theta\Theta}$ and $l_{\Theta\Phi}$. Now, let $I_{l_{\Theta\Theta}}$ and $I_{l_{\Phi\Phi}}$ be the limiting values of $-l_{\Theta\Theta}/N$ and $-l_{\Phi\Phi}/N$, respectively, under $h$ and the limiting values of $-l_{\Theta\Phi}/N$ and $-l_{\Phi\Theta}/N$, under $h$, are denoted by $I_{l_{\Theta\Phi}}$ and $I_{l_{\Phi\Theta}}$, respectively. Note that these limiting values are all evaluated at $\Theta_0$ and $\Phi_0$.

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

$$A = I_{l_{\Theta\Theta}} - I_{l_{\Theta\Phi}}I_{l_{\Phi\Phi}}^{-1}I_{l_{\Phi\Theta}}$$

and

$$B = V_{l_{\Theta\Theta}} - I_{l_{\Theta\Phi}}I_{l_{\Phi\Phi}}^{-1}V_{l_{\Phi\Theta}} - V_{l_{\Theta\Phi}}I_{l_{\Phi\Phi}}^{-1}I_{l_{\Phi\Theta}} + I_{l_{\Theta\Phi}}I_{l_{\Phi\Phi}}^{-1}V_{l_{\Phi\Phi}}V_{l_{\Phi\Theta}}^{-1}I_{l_{\Phi\Theta}}.$$

Here $V_{l_{\Theta\Theta}} = \lim_{N \to \infty} E_h[l_{\Theta}(\Theta_0, \Phi_0)l_{\Theta}(\Theta_0, \Phi_0)/N], V_{l_{\Theta\Phi}} = \lim_{N \to \infty} E_h[l_{\Theta}(\Theta_0, \Phi_0)l_{\Phi}(\Theta_0, \Phi_0)/N]$ and $V_{l_{\Phi\Phi}} = \lim_{N \to \infty} E_h[l_{\Phi}(\Theta_0, \Phi_0)l_{\Phi}(\Theta_0, \Phi_0)/N]$, where $E_h$ stands for the expectation evaluated under $h$.

Let $\hat{\Theta}$ be the MLE of $\Theta$ and $\hat{\Phi}$ and $\hat{B}$ be the empirical versions of $A$ and $B$. A direct application of Taylor’s expansion shows that the adjusted Wald statistic $N(\hat{\Theta} - \Theta_0)^T \hat{B}^{-1}(\hat{\Theta} - \Theta_0)$ has an asymptotic $\chi^2_p$ distribution for general $h_i, i = 1, 2, \ldots, N$, that have finite second moments. Here $\chi^2_p$ is denoted as a chi-squared distribution with $p$ degrees of freedom. Another asymptotically equivalent counterpart, the adjusted score statistic $N^{-1}[l_{\Phi}(\Theta_0, \Phi_0)] \hat{B}^{-1}(\Theta_0, \Phi(\Theta_0)) - l_{\Phi}(\Theta_0, \Phi(\Theta_0))$, where $\Phi(\Theta_0)$ and $\hat{B}(\Theta_0, \Phi(\Theta_0))$ are the constrained MLEs of $\Phi$ and $B$ given $\Theta_0$, respectively, has the same limiting $\chi^2_p$ 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 but not identically distributed random variables $Y_1, Y_2, \ldots, Y_N$. Under GLMs, the mean response, $\mu_i$, depends on the $p$ covariates $(x_{i0}, x_{i1}, \ldots, x_{ip-1}) = x_i^T$, by $\mu_i = g(\eta_i)$, where $\eta_i = x_i^T \gamma = y_0x_{i0} + y_1x_{i1} + \cdots + y_{p-1}x_{ip-1}$ is a linear predicator with the $p$ regression coefficients $(y_0, y_1, \ldots, y_{p-1}) = \gamma$, 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 $i$th 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 $\gamma_{j-1}$ is

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

where $\mu_i'$ is the first derivative of $\mu_i$ w.r.t. $\eta_i$. 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 conditions, the consistency of the MQL estimates under model misspecification depends only on $\gamma_0$ and $\mu_0'$.

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

$$A_n^{(uv)} = \lim_{N \to \infty} \frac{1}{N\sigma_0^2} \sum_{i=1}^{N} (\mu_i'(0))^2 (x_{i,p-u} - \sum_{j=1}^{p-w} \frac{|A_{nj}(0)|}{|A_n|} x_{i,j-1}) (x_{i,p-v} - \sum_{j=1}^{p-w} \frac{|A_{nj}(0)|}{|A_n|} x_{i,j-1})$$

and

$$B_n^{(uv)} = \lim_{N \to \infty} \frac{1}{N\sigma_0^4} \sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_i'(0))^2 (x_{i,p-u} - \sum_{j=1}^{p-w} \frac{|A_{nj}(0)|}{|A_n|} x_{i,j-1}) \times (x_{i,p-v} - \sum_{j=1}^{p-w} \frac{|A_{nj}(0)|}{|A_n|} x_{i,j-1}).$$

Here $|A_n|$ represents the determinant of the matrix $A_n$, where $A_n = W V_n^{-1} W$ with $W = (z_0, \ldots, z_{j-2}, z_{j-1}, z_j, \ldots, z_{p-w})$ and $V_n = \text{diag}((1/\mu_1'^2, (1/\mu_2'^2)^2, \ldots, (1/\mu_{N-1}'^2)^2)$ being a diagonal matrix of order $N$. On the other hand, $A_{nj}(u) = W V_n^{-1} W_{j}(u)$ and $A_{nj}(v) = W V_n^{-1} W_{j}(v)$ with $W_{j}(u) = (z_0, \ldots, z_{j-2}, z_{p-u}, z_j, \ldots, z_{p-w-1})$ and $W_{j}(v) = (z_0, \ldots, z_{j-2}, z_{p-v}, z_j, \ldots, z_{p-w-1})$ derived by the $j$th column $z_{j-1}$ of $W$ replaced by $z_{p-u}$ and $z_{p-v}$, respectively. Here $\sigma_0^2$ is the limit of the MLE of $\sigma^2$, $\hat{\sigma}^2$, that has the same interpretation of the limit of $\sum_{i=1}^{N} \text{Var}_h(Y_i)/N$. Note that the interpretation of $\sigma_0^2$ depends on $h$ and is, therefore, unknown.

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

$$A_n^{(uv)} = \lim_{N \to \infty} \frac{1}{N\sigma_0^2} \sum_{i=1}^{N} (\mu_i'(0))^2 (x_{i,p-u})(x_{i,p-v})$$

and

$$B_n^{(uv)} = \lim_{N \to \infty} \frac{1}{N\sigma_0^4} \sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_i'(0))^2 (x_{i,p-u})(x_{i,p-v}).$$
In applications, consistent estimates $\hat{A}_n$ and $\hat{B}_n$ of $A_n$ and $B_n$ can be obtained by $\text{Var}_h(Y_i)$ replaced by $(y_i - \hat{\mu})^2$ with $\hat{\mu}$ being the MLE of $\mu$ and other unknown quantities replaced by their respective empirical versions.

Let $\beta_0$ be the true value of $\beta$ and consider the null hypothesis $H_0 : \beta = \beta_0$. Let $\hat{\beta}$ be the MLE of $\beta$ based on the normal working model and let $\hat{\phi}(\beta_0)$ and $\hat{B}_n(\beta_0, \hat{\phi}(\beta_0))$ be the restricted MLEs of $\phi$ and $B_n$ given $\beta_0$. Here the vector of the nuisance parameters, $\phi$, contains the scale parameter $\sigma^2$ and some regression coefficients that are not to be tested. Under $H_0$, the adjusted Wald statistic $N(\hat{\beta} - \beta_0)^T A_n \hat{B}_n^{-1} A_n(\hat{\beta} - \beta_0)$ and the adjusted score statistic $N^{-1} \{l_\beta(\beta_0, \phi(\beta_0))\} \hat{B}_n^{-1}(\beta_0, \hat{\phi}(\beta_0)) \{l_\beta(\beta_0, \hat{\phi}(\beta_0))\}$ are asymptotically equivalent and have an asymptotic $\chi^2$ distribution as long as the second moments of the true underlying distributions exist. Note that $\hat{A}_n \hat{B}_n^{-1} \hat{A}_n$ is free of $\sigma^2$. Thus, with large samples, the effect of $\sigma^2$ is actually removed. Hence, $\sigma^2$ 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 $i$th 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

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

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 $A_n$ and $B_n$ show that the $(u, v)$, $u, v = 1, 2, \ldots, w$, components of the adjusting matrices $A_g$ and $B_g$ under the gamma working model are of the forms (for details, see the appendix):

$$A_{g(uv)} = \lim_{N \to \infty} \frac{r_0}{N} \sum_{i=1}^{N} \left( \frac{\mu_{i,0}}{\mu_i} \right)^2 \left( x_{i,p-u} - \sum_{j=1}^{p-w} \frac{|\Delta_{g(u)}|}{|\Delta_g|} x_{i,j-1} \right) \left( x_{i,p-v} - \sum_{j=1}^{p-w} \frac{|\Delta_{g(v)}|}{|\Delta_g|} x_{i,j-1} \right)$$

and

$$B_{g(uv)} = \lim_{N \to \infty} \frac{r_0^2}{N} \sum_{i=1}^{N} \text{Var}_h(Y_i) \left( \frac{\mu_{i,0}'}{\mu_i} \right)^2 \left( x_{i,p-u} - \sum_{j=1}^{p-w} \frac{|\Delta_{g(u)}|}{|\Delta_g|} x_{i,j-1} \right) \times \left( x_{i,p-v} - \sum_{j=1}^{p-w} \frac{|\Delta_{g(v)}|}{|\Delta_g|} x_{i,j-1} \right),$$

where $\Delta_g = W'V_g^{-1} W$, $\Delta_{g(u)} = W'V_g^{-1} W_{j(u)}$ and $\Delta_{g(v)} = W'V_g^{-1} W_{j(v)}$ with $V_g = \text{diag}((\mu_{1,0}/\mu_{1,0}')^2$, $(\mu_{2,0}/\mu_{2,0}')^2, \ldots, (\mu_{N,0}/\mu_{N,0}')^2)$. 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, \( A_g \) and \( B_g \) reduce to
\[
A_{g(a)} = \lim_{N \to \infty} \frac{r_0}{N} \sum_{i=1}^{N} \left( \frac{\mu_{i,0}'}{\mu_{i,0}} \right)^2 (x_{i,p-u})(x_{i,p-v})
\]
and
\[
B_{g(a)} = \lim_{N \to \infty} \frac{r_0^2}{N} \sum_{i=1}^{N} \frac{\text{Var}(Y_i)}{\mu_{i,0}^2} \left( \frac{\mu_{i,0}'}{\mu_{i,0}} \right)^2 (x_{i,p-u})(x_{i,p-v}).
\]

In application, consistent estimates \( \hat{A}_g \) and \( \hat{B}_g \) of \( A_g \) and \( B_g \) can be derived by replacing the unknown components in \( A_g \) and \( B_g \) by their respective empirical analogues, just as we dealt with \( \hat{A}_n \) and \( \hat{B}_n \). Note that \( \hat{A}_g \hat{B}_g^{-1} \hat{A}_g \) 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 \( N(\hat{\beta} - \beta_0)^\prime \hat{A}_g \hat{B}_g^{-1} \hat{A}_g (\hat{\beta} - \beta_0) \) and the resulting adjusted score statistic \( N^{-1} \{ \hat{Y}_p^\prime (\beta_0, \hat{\phi}(\beta_0)) \hat{B}_g^{-1} (\beta_0, \hat{\phi}(\beta_0))(\hat{Y}_p(\beta_0, \hat{\phi}(\beta_0))) \} \), under \( H_0 : \beta = \beta_0 \), are asymptotically distributed as \( \chi^2_n \) 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

- **Model 1**: \( \mu_i = \exp(\eta_i) \),
- **Model 2**: \( \mu_i = (2.5\eta_i + 2/3)^2 \),
- **Model 3**: \( \mu_i = \eta_i^2 \)

with the linear predictor \( \eta_i \) given by
\[
\eta_i = x_{i,0} + \gamma_1 x_{i,1} + \gamma_2 x_{i,2} \quad \text{for} \quad i = 1, 2, \ldots, N,
\]
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 \( \gamma_1 \) and \( \gamma_2 \) are considered as the parameters of interest. For simplicity, let \( \gamma' = (\gamma_1, \gamma_2) \) and \( \gamma_1' = (1.0, 1.0) \). We test the null hypothesis \( H_0 : \gamma = \gamma_0 \) and the two alternative hypotheses \( H_A : \gamma' = (0.4, 1.0) \) and \( \gamma' = (0.7, 1.3) \), 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 \( \lambda \) and the scale parameter \( k \), \( \text{W}(k, \lambda) \), has a simple relationship between the second central moment and the first moment, that is, \( \text{Var}(Y) = a\mu^2 \), where \( a > 0 \) is a function of the shape parameter \( \lambda \). For example, when \( \lambda = 1 \) and \( k = \mu \), \( \text{Vr}(Y) = \mu^2 \). Similarly, an inverse Gaussian distribution with the mean \( \mu \) and the shape parameter \( \lambda \), \( \text{IG}(\mu, \lambda) \), has a variance proportional to the cubic of its mean value, that is, \( \text{Var}(Y) = \mu^3 / \lambda \). On the other hand, a non-central chi-squared distribution with \( v \) degrees of freedom and a non-centrality parameter \( \mu - v > 0 \), \( \chi^2_v(\mu - v) \), has a mean value of \( \mu \) and a variance of \( 2(2\mu - v) \), so that \( \chi^2_v(\mu - v) \) 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 0.3\(N \) 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 \( \mu_i \). Then, its two asymptotically equivalent test statistics, the Wald test statistic and the score test statistic, are defined by

\[
Q_L = 2[l(\hat{\mu}, \hat{\phi}) - l(\mu_0, \hat{\phi}(\mu_0))].
\]

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

\[
Q_W = N(\hat{\mu} - \mu_0)'A(\hat{\mu} - \mu_0)
\]

and

\[
Q_S = N^{-1}\{l^l_y(\mu_0, \hat{\phi}(\mu_0))A^{-1}(\mu_0, \hat{\phi}(\mu_0))l_y(\mu_0, \hat{\phi}(\mu_0))\},
\]

respectively. Here all notation definitions are given as in the previous sections. Each of the test statistics \( Q_L, Q_W \) and \( Q_S \), under the null hypothesis \( H_0 : \gamma = \gamma_0 \), has an asymptotic chi-squared distribution with degrees of freedom equal to the dimension of \( \gamma \). 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 \( Q_L, Q_W \) 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 \( \chi^2_2 \). More discussions about the test statistics \( Q_L, Q_W \) and \( Q_S \) can be found in [28, Section 9.3].

The simulation performance are carried out for 3000 simulation runs with the \( x_{i1} \)'s and \( x_{i2} \)'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 \( AW_\alpha, AS_\alpha, L_\alpha, W_\alpha \) and \( S_\alpha \), 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 \( H_0 : \gamma = \gamma_0 \) 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.

The average of the 3000 \( \hat{\gamma} \) values and their sample covariance matrix are termed as \( \bar{\gamma} \) 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 \( \hat{\gamma} \), namely \( \hat{\gamma}^{-1}/N \), denoted
Table 1. Model 1: $\mu_i = \exp(\theta_i), i = 1, 2, \ldots, N.$

<table>
<thead>
<tr>
<th>Working model mean($\hat{\gamma}$)</th>
<th>$S^2$</th>
<th>Var$_A$($\hat{\gamma}$)</th>
<th>Var($\hat{\gamma}$)</th>
<th>AW$\alpha$</th>
<th>AWcp</th>
<th>AS$\alpha$</th>
<th>AScp</th>
<th>LA</th>
<th>Lcp</th>
<th>W$\alpha$</th>
<th>Wcp</th>
<th>Sa</th>
<th>Scp</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>$N = 450$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_0 : \gamma = (1.0, 1.0)$</td>
<td>Normal</td>
<td>$[1.0030, 1.0024]$</td>
<td>[0.0223, 0.0054, 0.0242]</td>
<td>[0.0219, 0.0040, 0.0216]</td>
<td>[0.0161, 0.0001, 0.0161]</td>
<td>0.0643</td>
<td>0.9357</td>
<td>0.0587</td>
<td>0.9413</td>
<td>0.1190</td>
<td>0.8810</td>
<td>0.1223</td>
<td>0.8777</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>$[0.9992, 0.9998]$</td>
<td>[0.0143, 0.0005, 0.0150]</td>
<td>[0.0145, 0.0004, 0.0143]</td>
<td>[0.0165, 0.0001, 0.0165]</td>
<td>0.0597</td>
<td>0.9403</td>
<td>0.0407</td>
<td>0.9530</td>
<td>0.0307</td>
<td>0.9693</td>
<td>0.0290</td>
<td>0.9710</td>
</tr>
<tr>
<td>$H_A : \gamma = (0.4, 1.0)$</td>
<td>Normal</td>
<td>$[0.3980, 0.0038]$</td>
<td>[0.0198, 0.0018, 0.0203]</td>
<td>[0.0193, 0.0014, 0.0194]</td>
<td>[0.0152, 0.0001, 0.0177]</td>
<td>0.9753</td>
<td>0.9377</td>
<td>0.9673</td>
<td>0.9413</td>
<td>0.9860</td>
<td>0.9133</td>
<td>0.9883</td>
<td>0.9137</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>$[0.3980, 0.0038]$</td>
<td>[0.0154, 0.0003, 0.0158]</td>
<td>[0.0154, 0.0001, 0.0154]</td>
<td>[0.0184, 0.0001, 0.0184]</td>
<td>0.9950</td>
<td>0.9350</td>
<td>0.9474</td>
<td>0.9473</td>
<td>0.9923</td>
<td>0.9690</td>
<td>0.9927</td>
<td>0.9690</td>
</tr>
<tr>
<td>$H_A : \gamma = (0.7, 1.3)$</td>
<td>Normal</td>
<td>$[0.7027, 1.3032]$</td>
<td>[0.0227, 0.0049, 0.0248]</td>
<td>[0.0219, 0.0036, 0.0227]</td>
<td>[0.0147, 0.0001, 0.0181]</td>
<td>0.8280</td>
<td>0.9393</td>
<td>0.8150</td>
<td>0.9443</td>
<td>0.8573</td>
<td>0.8803</td>
<td>0.8587</td>
<td>0.8730</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>$[0.7005, 1.2987]$</td>
<td>[0.0145, 0.0005, 0.0151]</td>
<td>[0.0145, 0.0003, 0.0145]</td>
<td>[0.0166, 0.0001, 0.0166]</td>
<td>0.8970</td>
<td>0.9453</td>
<td>0.8860</td>
<td>0.9530</td>
<td>0.8647</td>
<td>0.8637</td>
<td>0.9653</td>
<td>0.8573</td>
</tr>
<tr>
<td><strong>$N = 900$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_0 : \gamma = (1.0, 1.0)$</td>
<td>Normal</td>
<td>$[1.0033, 1.0003]$</td>
<td>[0.0120, 0.0026, 0.0112]</td>
<td>[0.0113, 0.0024, 0.0109]</td>
<td>[0.0081, 0.0000, 0.0080]</td>
<td>0.0603</td>
<td>0.9397</td>
<td>0.0593</td>
<td>0.9407</td>
<td>0.1187</td>
<td>0.8813</td>
<td>0.1203</td>
<td>0.8797</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>$[1.0012, 0.9995]$</td>
<td>[0.0074, 0.0001, 0.0072]</td>
<td>[0.0073, 0.0002, 0.0072]</td>
<td>[0.0083, 0.0000, 0.0082]</td>
<td>0.0593</td>
<td>0.9407</td>
<td>0.0517</td>
<td>0.9483</td>
<td>0.0357</td>
<td>0.9643</td>
<td>0.0357</td>
<td>0.9643</td>
</tr>
<tr>
<td>$H_A : \gamma = (0.4, 1.0)$</td>
<td>Normal</td>
<td>$[0.3992, 1.0009]$</td>
<td>[0.0102, 0.0010, 0.0098]</td>
<td>[0.0099, 0.0008, 0.0096]</td>
<td>[0.0076, 0.0000, 0.0088]</td>
<td>0.9997</td>
<td>0.9403</td>
<td>0.9993</td>
<td>0.9437</td>
<td>0.9997</td>
<td>0.9120</td>
<td>1.0000</td>
<td>0.9100</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>$[0.3984, 0.9997]$</td>
<td>[0.0082, 0.0002, 0.0076]</td>
<td>[0.0078, 0.0001, 0.0077]</td>
<td>[0.0092, 0.0000, 0.0091]</td>
<td>1.0000</td>
<td>0.9407</td>
<td>1.0000</td>
<td>0.9477</td>
<td>1.0000</td>
<td>0.9647</td>
<td>1.0000</td>
<td>0.9633</td>
</tr>
<tr>
<td>$H_A : \gamma = (0.7, 1.3)$</td>
<td>Normal</td>
<td>$[0.7019, 1.3011]$</td>
<td>[0.0118, 0.0023, 0.0115]</td>
<td>[0.0113, 0.0022, 0.0113]</td>
<td>[0.0073, 0.0000, 0.0090]</td>
<td>0.9853</td>
<td>0.9400</td>
<td>0.9847</td>
<td>0.9437</td>
<td>0.9900</td>
<td>0.8793</td>
<td>0.9893</td>
<td>0.8763</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>$[0.7008, 1.3005]$</td>
<td>[0.0073, 0.0003, 0.0070]</td>
<td>[0.0073, 0.0001, 0.0072]</td>
<td>[0.0083, 0.0000, 0.0082]</td>
<td>0.9953</td>
<td>0.9453</td>
<td>0.9953</td>
<td>0.9510</td>
<td>0.9937</td>
<td>0.9647</td>
<td>0.9937</td>
<td>0.9643</td>
</tr>
</tbody>
</table>
Table 1. Continued

<table>
<thead>
<tr>
<th>Working model mean($\hat{\gamma}$)</th>
<th>$S^2$</th>
<th>Var$_A$($\hat{\gamma}$)</th>
<th>Var($\hat{\gamma}$)</th>
<th>AW$\alpha$</th>
<th>AW$\alpha$</th>
<th>AS$\alpha$</th>
<th>AS$\alpha$</th>
<th>L$\alpha$</th>
<th>L$\alpha$</th>
<th>W$\alpha$</th>
<th>W$\alpha$</th>
<th>s$\alpha$</th>
<th>s$\alpha$</th>
<th>S$\alpha$</th>
<th>S$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$ = 1350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_0 : \gamma = (1.0, 1.0)$</td>
<td>Normal</td>
<td>[1.0022, 1.0015]</td>
<td>[0.0076, 0.0073]</td>
<td>[0.0073, 0.0073]</td>
<td>[0.0054, 0.0054]</td>
<td>0.0560</td>
<td>0.9440</td>
<td>0.0543</td>
<td>0.9457</td>
<td>0.1123</td>
<td>0.8877</td>
<td>0.1130</td>
<td>0.8870</td>
<td>0.1100</td>
<td>0.8900</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>[0.0013, 0.0014]</td>
<td>[0.0049, 0.0047]</td>
<td>[0.0048, 0.0048]</td>
<td>[0.0055, 0.0055]</td>
<td>0.0527</td>
<td>0.9473</td>
<td>0.0450</td>
<td>0.9550</td>
<td>0.0320</td>
<td>0.9680</td>
<td>0.0327</td>
<td>0.9673</td>
<td>0.0323</td>
<td>0.9677</td>
</tr>
<tr>
<td>$H_A : \gamma = (0.4, 1.0)$</td>
<td>Normal</td>
<td>[0.4003, 1.0033]</td>
<td>[0.0068, 0.0068]</td>
<td>[0.0064, 0.0065]</td>
<td>[0.0050, 0.0050]</td>
<td>1.0000</td>
<td>0.9367</td>
<td>1.0000</td>
<td>0.9393</td>
<td>1.0000</td>
<td>0.9093</td>
<td>1.0000</td>
<td>0.9083</td>
<td>1.0000</td>
<td>0.9093</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>[0.4003, 1.0023]</td>
<td>[0.0054, 0.0054]</td>
<td>[0.0052, 0.0052]</td>
<td>[0.0061, 0.0061]</td>
<td>1.0000</td>
<td>0.9410</td>
<td>1.0000</td>
<td>0.9460</td>
<td>1.0000</td>
<td>0.9680</td>
<td>1.0000</td>
<td>0.9670</td>
<td>1.0000</td>
<td>0.9693</td>
</tr>
<tr>
<td>$H_A : \gamma = (0.7, 1.3)$</td>
<td>Normal</td>
<td>[0.7002, 1.3016]</td>
<td>[0.0075, 0.0078]</td>
<td>[0.0073, 0.0076]</td>
<td>[0.0049, 0.0060]</td>
<td>0.9983</td>
<td>0.9373</td>
<td>0.9983</td>
<td>0.9430</td>
<td>0.9983</td>
<td>0.8793</td>
<td>0.9983</td>
<td>0.8790</td>
<td>0.9983</td>
<td>0.8813</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>[0.6993, 1.3005]</td>
<td>[0.0049, 0.0048]</td>
<td>[0.0048, 0.0049]</td>
<td>[0.0055, 0.0055]</td>
<td>0.9993</td>
<td>0.9427</td>
<td>0.9993</td>
<td>0.9463</td>
<td>0.9993</td>
<td>0.9663</td>
<td>0.9993</td>
<td>0.9623</td>
<td>0.9993</td>
<td>0.9637</td>
</tr>
</tbody>
</table>
### Table 2. Model 2: $\mu_i = (2.5\eta_i + 2/3)^3$, $i = 1, 2, \ldots, N.$

<table>
<thead>
<tr>
<th>Working model</th>
<th>mean($\hat{\gamma}$)</th>
<th>$S^2$</th>
<th>Var($\hat{\gamma}$)</th>
<th>Var($\hat{\rho}$)</th>
<th>AWa</th>
<th>AWcp</th>
<th>ASa</th>
<th>AScp</th>
<th>LA</th>
<th>Lcp</th>
<th>Wa</th>
<th>Wcp</th>
<th>Sa</th>
<th>Scp</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N = 450$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_0 : \gamma = (1.0, 1.0)$</td>
<td>Normal</td>
<td>[1.0094, 1.0093]</td>
<td>[0.0289, 0.0271]</td>
<td>[0.0269, 0.0268]</td>
<td>[0.0152, 0.0152]</td>
<td>0.0910</td>
<td>0.0900</td>
<td>0.0897</td>
<td>0.0913</td>
<td>0.1533</td>
<td>0.8467</td>
<td>0.1560</td>
<td>0.8440</td>
<td>0.1507</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>[0.9997, 0.9998]</td>
<td>[0.0068, 0.0065]</td>
<td>[0.0068, 0.0067]</td>
<td>[0.0049, 0.0049]</td>
<td>0.0657</td>
<td>0.9343</td>
<td>0.0510</td>
<td>0.9490</td>
<td>0.1047</td>
<td>0.8953</td>
<td>0.1067</td>
<td>0.8933</td>
<td>0.1307</td>
</tr>
<tr>
<td>$H_A : \gamma = (0.7, 1.3)$</td>
<td>Normal</td>
<td>[0.4030, 1.0058]</td>
<td>[0.0114, 0.0111]</td>
<td>[0.0107, 0.0112]</td>
<td>[0.0067, 0.0081]</td>
<td>0.9983</td>
<td>0.9230</td>
<td>0.9967</td>
<td>0.9247</td>
<td>0.9980</td>
<td>0.8730</td>
<td>0.9987</td>
<td>0.8723</td>
<td>0.9980</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>[0.3998, 0.9998]</td>
<td>[0.0038, 0.0041]</td>
<td>[0.0038, 0.0043]</td>
<td>[0.0033, 0.0034]</td>
<td>1.0000</td>
<td>0.9443</td>
<td>1.0000</td>
<td>0.9500</td>
<td>1.0000</td>
<td>0.9217</td>
<td>1.0000</td>
<td>0.9223</td>
<td>1.0000</td>
</tr>
<tr>
<td>$N = 900$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_0 : \gamma = (1.0, 1.0)$</td>
<td>Normal</td>
<td>[1.0067, 1.0107]</td>
<td>[0.0661, 0.0674]</td>
<td>[0.0566, 0.0531]</td>
<td>[0.0302, 0.0299]</td>
<td>0.1123</td>
<td>0.8877</td>
<td>0.1090</td>
<td>0.8910</td>
<td>0.1723</td>
<td>0.8277</td>
<td>0.1763</td>
<td>0.8273</td>
<td>0.1650</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>[0.9903, 0.9946]</td>
<td>[0.0144, 0.0141]</td>
<td>[0.0133, 0.0132]</td>
<td>[0.0099, 0.0099]</td>
<td>0.0877</td>
<td>0.9123</td>
<td>0.0640</td>
<td>0.9360</td>
<td>0.1190</td>
<td>0.8810</td>
<td>0.1200</td>
<td>0.8800</td>
<td>0.1167</td>
</tr>
<tr>
<td>$H_A : \gamma = (0.4, 1.0)$</td>
<td>Normal</td>
<td>[0.3974, 1.0059]</td>
<td>[0.0221, 0.0247]</td>
<td>[0.0209, 0.0225]</td>
<td>[0.0132, 0.0160]</td>
<td>0.9753</td>
<td>0.9027</td>
<td>0.9690</td>
<td>0.9103</td>
<td>0.9767</td>
<td>0.8677</td>
<td>0.9783</td>
<td>0.8650</td>
<td>0.9737</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>[0.3953, 0.9964]</td>
<td>[0.0081, 0.0090]</td>
<td>[0.0076, 0.0085]</td>
<td>[0.0066, 0.0068]</td>
<td>1.0000</td>
<td>0.9257</td>
<td>1.0000</td>
<td>0.9457</td>
<td>1.0000</td>
<td>0.9093</td>
<td>1.0000</td>
<td>0.9080</td>
<td>1.0000</td>
</tr>
<tr>
<td>$H_A : \gamma = (0.7, 1.3)$</td>
<td>Normal</td>
<td>[0.7022, 1.3157]</td>
<td>[0.0637, 0.0785]</td>
<td>[0.0553, 0.0605]</td>
<td>[0.0283, 0.0350]</td>
<td>0.6840</td>
<td>0.8790</td>
<td>0.6817</td>
<td>0.8853</td>
<td>0.7183</td>
<td>0.8220</td>
<td>0.7190</td>
<td>0.8197</td>
<td>0.7127</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>[0.6915, 1.2931]</td>
<td>[0.0131, 0.0149]</td>
<td>[0.0123, 0.0140]</td>
<td>[0.0096, 0.0100]</td>
<td>0.9270</td>
<td>0.9137</td>
<td>0.9107</td>
<td>0.9337</td>
<td>0.9530</td>
<td>0.8760</td>
<td>0.9557</td>
<td>0.8750</td>
<td>0.9497</td>
</tr>
<tr>
<td>$H_A : \gamma = (0.7, 1.3)$</td>
<td>Normal</td>
<td>[0.9997, 0.9998]</td>
<td>[0.0068, 0.0065]</td>
<td>[0.0068, 0.0067]</td>
<td>[0.0049, 0.0049]</td>
<td>0.9983</td>
<td>0.9230</td>
<td>0.9967</td>
<td>0.9247</td>
<td>0.9980</td>
<td>0.8730</td>
<td>0.9987</td>
<td>0.8723</td>
<td>0.9980</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>[0.6998, 1.2995]</td>
<td>[0.0063, 0.0071]</td>
<td>[0.0063, 0.0072]</td>
<td>[0.0048, 0.0049]</td>
<td>0.9983</td>
<td>0.9313</td>
<td>0.9973</td>
<td>0.9497</td>
<td>0.9987</td>
<td>0.8933</td>
<td>0.9990</td>
<td>0.8917</td>
<td>0.9987</td>
</tr>
</tbody>
</table>
Table 2. Continued

<table>
<thead>
<tr>
<th>Working model</th>
<th>mean((\hat{\gamma}))</th>
<th>(S^2)</th>
<th>Var((\hat{\gamma}))</th>
<th>AW((\alpha))</th>
<th>AWcp</th>
<th>AS((\alpha))</th>
<th>AScp</th>
<th>La</th>
<th>Lcp</th>
<th>Wa</th>
<th>Wcp</th>
<th>S((\alpha))</th>
<th>Scp</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H_0: \gamma = (1.0, 1.0))</td>
<td>Normal</td>
<td>[1.0010, 1.0048]</td>
<td>[0.0170, 0.0083, 0.0173]</td>
<td>[0.0169, 0.0081, 0.0174]</td>
<td>[0.0099, 0.0005, 0.0100]</td>
<td>0.9039, 0.9097, 0.0897, 0.9103, 0.1560, 0.8440, 0.1563, 0.8437, 0.1513, 0.8487</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>[0.9982, 1.0005]</td>
<td>[0.0045, 0.0002, 0.0044]</td>
<td>[0.0045, 0.0001, 0.0045]</td>
<td>[0.0033, 0.0001, 0.0033]</td>
<td>0.0610, 0.9390, 0.0500, 0.9500, 0.1183, 0.8817, 0.1167, 0.8833, 0.1157, 0.8843</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(H_A: \gamma = (0.4, 1.0))</td>
<td>Normal</td>
<td>[0.3990, 1.0034]</td>
<td>[0.0070, 0.0020, 0.0073]</td>
<td>[0.0070, 0.0020, 0.0074]</td>
<td>[0.0044, 0.0001, 0.0053]</td>
<td>1.0000, 0.9313, 1.0000, 0.9340, 1.0000, 0.8737, 1.0000, 0.8727, 1.0000, 0.8743</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>[0.3994, 1.0006]</td>
<td>[0.0025, 0.0001, 0.0028]</td>
<td>[0.0026, 0.0001, 0.0029]</td>
<td>[0.0022, 0.0000, 0.0023]</td>
<td>1.0000, 0.9460, 1.0000, 0.9530, 1.0000, 0.9107, 1.0000, 0.9113, 1.0000, 0.9103</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(H_A: \gamma = (0.7, 1.3))</td>
<td>Normal</td>
<td>[0.7001, 1.3077]</td>
<td>[0.0172, 0.0078, 0.0193]</td>
<td>[0.0172, 0.0077, 0.0193]</td>
<td>[0.0095, 0.0005, 0.0116]</td>
<td>0.9850, 0.9123, 0.9870, 0.9160, 0.9880, 0.8430, 0.9863, 0.8400, 0.9877, 0.8450</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>[0.6986, 1.3008]</td>
<td>[0.0042, 0.0002, 0.0048]</td>
<td>[0.0042, 0.0001, 0.0049]</td>
<td>[0.0032, 0.0001, 0.0033]</td>
<td>0.9997, 0.9417, 0.9997, 0.9497, 1.0000, 0.8907, 1.0000, 0.8917, 1.0000, 0.8897</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td>([\gamma_1, \gamma_2])</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(N = 1350\)
<table>
<thead>
<tr>
<th>Working model</th>
<th>mean($\hat{\gamma}$)</th>
<th>$S^2$</th>
<th>Var($\hat{\gamma}$)</th>
<th>Var($\hat{\phi}$)</th>
<th>$\text{AW}_\alpha$</th>
<th>$\text{AW}_{cp}$</th>
<th>$\text{AS}_\alpha$</th>
<th>$\text{AS}_{cp}$</th>
<th>$\text{L}_\alpha$</th>
<th>$\text{L}_{cp}$</th>
<th>$\text{W}_\alpha$</th>
<th>$\text{W}_{cp}$</th>
<th>$\text{Scp}$</th>
<th>$\text{S}_\alpha$</th>
<th>$\text{Scp}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0 : \gamma = (1, 0)$</td>
<td>Normal [1.0004, 1.0010]</td>
<td>0.0128</td>
<td>0.0019</td>
<td>0.0121</td>
<td>0.0020</td>
<td>0.0117</td>
<td>0.0020</td>
<td>0.0102</td>
<td>0.0044</td>
<td>0.0101</td>
<td>0.0567</td>
<td>0.9433</td>
<td>0.0553</td>
<td>0.9447</td>
<td>0.0743</td>
</tr>
<tr>
<td></td>
<td>Gamma [1.0003, 1.0023]</td>
<td>0.0087</td>
<td>−0.0003</td>
<td>0.0083</td>
<td>−0.0002</td>
<td>0.0082</td>
<td>−0.0002</td>
<td>0.0096</td>
<td>−0.0005</td>
<td>0.0095</td>
<td>0.0627</td>
<td>0.9373</td>
<td>0.0503</td>
<td>0.9497</td>
<td>0.0367</td>
</tr>
<tr>
<td>$H_A : \gamma = (0.4, 1, 0)$</td>
<td>Normal [0.4011, 1.0015]</td>
<td>0.0095</td>
<td>0.0008</td>
<td>0.0092</td>
<td>0.0008</td>
<td>0.0089</td>
<td>0.0008</td>
<td>0.0077</td>
<td>0.0002</td>
<td>0.0082</td>
<td>1.0000</td>
<td>0.9477</td>
<td>1.0000</td>
<td>0.9480</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>Gamma [0.4029, 1.0022]</td>
<td>0.0069</td>
<td>−0.0002</td>
<td>0.0067</td>
<td>−0.0001</td>
<td>0.0069</td>
<td>−0.0001</td>
<td>0.0077</td>
<td>−0.0002</td>
<td>0.0080</td>
<td>1.0000</td>
<td>0.9383</td>
<td>1.0000</td>
<td>0.9513</td>
<td>1.0000</td>
</tr>
<tr>
<td>$H_A : \gamma = (0.7, 1, 0)$</td>
<td>Normal [0.7019, 1.3034]</td>
<td>0.0130</td>
<td>0.0017</td>
<td>0.0125</td>
<td>0.0019</td>
<td>0.0115</td>
<td>0.0019</td>
<td>0.0100</td>
<td>0.0004</td>
<td>0.0107</td>
<td>0.9743</td>
<td>0.9450</td>
<td>0.9747</td>
<td>0.9487</td>
<td>0.9783</td>
</tr>
<tr>
<td></td>
<td>Gamma [0.7015, 1.3033]</td>
<td>0.0081</td>
<td>0.0000</td>
<td>0.0082</td>
<td>0.0000</td>
<td>0.0082</td>
<td>0.0000</td>
<td>0.0093</td>
<td>−0.0005</td>
<td>0.0096</td>
<td>0.9897</td>
<td>0.9433</td>
<td>0.9910</td>
<td>0.9503</td>
<td>0.9863</td>
</tr>
</tbody>
</table>
Table 3. Continued

<table>
<thead>
<tr>
<th>Working model</th>
<th>mean((\hat{y}))</th>
<th>(S^2)</th>
<th>Var_A((\hat{y}))</th>
<th>Var((\hat{y}))</th>
<th>AWa</th>
<th>AWcp</th>
<th>ASa</th>
<th>AScp</th>
<th>La</th>
<th>Lcp</th>
<th>Wa</th>
<th>Wcp</th>
<th>Sa</th>
<th>Scp</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H_0: \gamma = (1.0, 1.0))</td>
<td>Normal</td>
<td>[1.0025]</td>
<td>[0.0082 0.0013]</td>
<td>[0.0080 0.0013]</td>
<td>[0.0068 0.0002]</td>
<td>0.0627 0.9373 0.0603 0.9397 0.0860 0.9140 0.0860 0.9140 0.0857 0.9143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>[1.0007]</td>
<td>[0.0058 -0.0001]</td>
<td>[0.0055 -0.0001]</td>
<td>[0.0054 -0.0001]</td>
<td>0.0580 0.9420 0.0507 0.9493 0.0307 0.9693 0.0340 0.9660 0.0303 0.9697</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(H_A: \gamma = (0.4, 1.0))</td>
<td>Normal</td>
<td>[0.4021]</td>
<td>[0.0061 0.0005]</td>
<td>[0.0061 0.0005]</td>
<td>[0.0051 0.0001]</td>
<td>1.0000 0.9423 1.0000 0.9427 1.0000 0.9243 1.0000 0.9243 1.0000 0.9253</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>[0.4028]</td>
<td>[0.0046 0.0000]</td>
<td>[0.0045 -0.0001]</td>
<td>[0.0045 -0.0001]</td>
<td>1.0000 0.9380 1.0000 0.9447 1.0000 0.9660 1.0000 0.9650 1.0000 0.9643</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(H_A: \gamma = (0.7, 1.3))</td>
<td>Normal</td>
<td>[0.7009]</td>
<td>[0.0085 0.0009]</td>
<td>[0.0082 0.0012]</td>
<td>[0.0066 0.0002]</td>
<td>0.9960 0.9453 0.9957 0.9473 0.9950 0.9157 0.9950 0.9150 0.9947 0.9187</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>[0.7007]</td>
<td>[0.0056 -0.0002]</td>
<td>[0.0055 -0.0001]</td>
<td>[0.0055 -0.0001]</td>
<td>0.9990 0.9480 0.9993 0.9513 0.9983 0.9707 0.9987 0.9687 0.9987 0.9710</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(N = 1350\)
by $\text{Var}(\hat{\gamma})$ and the average of the adjusted covariance matrix estimate of $\hat{\gamma}$, namely $\hat{A}^{-1}\hat{B}\hat{A}^{-1}/N$, denoted by $\text{Var}_A(\hat{\gamma})$ 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 $\sigma^2$ and $r$, the non-regression parameters $\sigma^2$ 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, $\text{Var}_A(\hat{\gamma})$, are nearly equivalent to the sample covariance matrix of $\hat{\gamma}$, $S^2$, whereas the averages of the unadjusted covariance matrix estimates, $\text{Var}(\hat{\gamma})$, 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 $Q_L$, $Q_W$ 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 $AW_\alpha$ and $AS_\alpha$ are more close to the nominal significance level 0.05, in contrast with the values of $L_\alpha$, $W_\alpha$ and $S_\alpha$.

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 $AW_\alpha$ and $AS_\alpha$ 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 $Q_L$, $Q_W$ 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 $H_A : \gamma^* = (0.7, 1.3)$, 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, $\text{Var}(\hat{\gamma})$, 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 $\sigma^2$ 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 $\sigma^2$ 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

Appendix

Here some details regarding the quantities required to calculate the adjusting matrices $A_n$ and $B_n$ of $A_nB_n^{-1}A_n$ that correct the normal regression model are provided.

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

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

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

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

respectively. $I_{\phi\phi}$ is asymptotically expressed by

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

where $i_{\phi\phi} = -\sigma_0^2 I_{\beta\beta}^2$.

For simplicity of notation, let $\Delta_n$ be the $(p - w) \times (p - w)$ matrix with the $j$th row as $(\sum_{i=1}^{N} (\mu_i)'^2 x_{i,j-1}x_{i,0}, \ldots, \sum_{i=1}^{N} (\mu_i)'^2 x_{i,j-1}x_{i,p-w-1})$ for $j = 1, 2, \ldots, p - w$. Then, $I_{\phi\phi}$ is approximately written in the form

$$I_{\phi\phi} \approx \frac{1}{N\sigma_0^2} \begin{bmatrix} i_{\phi\phi} & 0 \\ 0 & \Delta_n \end{bmatrix},$$

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

$$I_{\phi\phi}^{-1} \approx \frac{N\sigma_0^2}{\|\Delta_n\|} \begin{bmatrix} \Delta_n & 0 & \cdots & 0 \\ 0 & R_{1,1} & \cdots & R_{p-w,1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & R_{1,p-w} & \cdots & R_{p-w,p-w} \end{bmatrix},$$

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


According to Equation (1), the \( \Delta_n \) are as follows:

\[
\begin{bmatrix}
\sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_{i,0}^{\prime})^2 x_{i,p-1}^2 & \ldots & \sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_{i,0}^{\prime})^2 x_{i,p-w}^2 \\
\vdots & \ddots & \vdots \\
\sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_{i,0}^{\prime})^2 x_{i,p-w}^2 & \ldots & \sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_{i,0}^{\prime})^2 x_{i,p-w}^2
\end{bmatrix}
\]

\[
\begin{bmatrix}
0 & \sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_{i,0}^{\prime})^2 x_{i,p-1}^2 & \ldots & 0 \\
0 & \sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_{i,0}^{\prime})^2 x_{i,p-1}^2 & \ldots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_{i,0}^{\prime})^2 x_{i,p-w}^2 & \ldots & 0
\end{bmatrix}
\]

and

\[
\begin{bmatrix}
\varphi_1 \varphi_0 & \varphi_1 \varphi_0^2 & \ldots & \varphi_1 \varphi_{p-w-1} \\
\varphi_0 \varphi_0^2 & \sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_{i,0}^{\prime})^2 x_{i,0}^2 & \ldots & \sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_{i,0}^{\prime})^2 x_{i,0} x_{i,p-w-1} \\
\vdots & \ddots & \ddots & \vdots \\
\varphi_{p-w-1} \varphi_0^2 & \sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_{i,0}^{\prime})^2 x_{i,p-w-1} x_{i,0} & \ldots & \sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_{i,0}^{\prime})^2 x_{i,p-w} x_{i,p-w-1}
\end{bmatrix}
\]

respectively. Here \( \varphi_1 = \sigma^2_0 E_h[^{\sigma_2} \sigma_0 \sigma_1 \sigma_0, \sigma_0 \sigma_1 \sigma_{p-w}] \) and \( \varphi_{p-w} = \sigma^2_0 E_h[^{\sigma_{p-w}} \sigma_{p-w}, \sigma_{p-w}, \sigma_{p-w}] \).

According to Equation (1), the \((u, v)\) entries of \( A_n \) are derived as follows:

\[
A_n(uv) = I_{h\beta \sigma} - I_{h\beta \sigma} \Phi T_{\h_0}^\prime \Phi T_{\h_0}^\prime \Phi T_{\h_0}^\prime + I_{h\beta \sigma} \Phi T_{\h_0}^\prime \Phi T_{\h_0}^\prime \Phi T_{\h_0}^\prime \Phi T_{\h_0}^\prime
\]

\[
= \lim_{N \to \infty} \frac{1}{N \sigma_0^2} \left\{ \sum_{i=1}^{N} (\mu_{i,0}^{\prime})^2 x_{i,p-u} x_{i,p-v} - \frac{1}{|A_n|} \sum_{i=1}^{N} (\mu_{i,0}^{\prime})^2 \left( x_{i,p-v} \sum_{j=1}^{p-w} x_{i,j-1} \right) \right\}
\]

\[
+ \frac{1}{|A_n|} \sum_{i=1}^{N} (\mu_{i,0}^{\prime})^2 \left( \sum_{j=1}^{p-w} x_{i,j-1} \right) \left( \sum_{j=1}^{p-w} x_{i,j-1} \right)
\]

\[
= \lim_{N \to \infty} \frac{1}{N \sigma_0^2} \left\{ \sum_{i=1}^{N} (\mu_{i,0}^{\prime})^2 \left( x_{i,p-u} \sum_{j=1}^{p-w} \frac{A_{n}(u) |A_n|}{|A_n|} x_{i,j-1} \right) \right\}
\]

\[
= \sum_{m=1}^{p-w} (R_{j,m} \sum_{i=1}^{N} (\mu_{i,0}^{\prime})^2 x_{i,m-1} x_{i,p-u})
\]

where \( |A_n| = \sum_{m=1}^{p-w} (R_{j,m} \sum_{i=1}^{N} (\mu_{i,0}^{\prime})^2 x_{i,m-1} x_{i,p-u}) \).
According to Equation (2), the \((u, v)\) entries of \(B_n\) are derived as follows:

\[
B_n(u, v) = V_h \beta_u \beta_v - V_h \beta_u \beta_v I_h \beta_v - I_h \beta_u \beta_v V_h \beta_v I_h \beta_v + I_h \beta_u \beta_v V_h \beta_v I_h \beta_v - V_h \beta_u \beta_v I_h \beta_v - I_h \beta_u \beta_v V_h \beta_v I_h \beta_v + I_h \beta_u \beta_v I_h \beta_v V_h \beta_v I_h \beta_v + I_h \beta_u \beta_v I_h \beta_v V_h \beta_v I_h \beta_v.
\]

\[
= \lim_{N \to \infty} \frac{1}{N \sigma_0^2} \sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_{i,0}')^2 \left( x_{i, p-u} - \sum_{j=1}^{p-w} x_{i, j-1} | \Delta_{n(i)} \right)
\]

\[
+ \frac{1}{|\Delta_n|^2} \sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_{i,0}')^2 \left( \sum_{j=1}^{p-w} x_{i, j-1} | \Delta_{n(i)} \right) \left( \sum_{j=1}^{p-w} x_{i, j-1} | \Delta_{n(i)} \right)
\]

\[
= \lim_{N \to \infty} \frac{1}{N \sigma_0^2} \sum_{i=1}^{N} \text{Var}_h(Y_i)(\mu_{i,0}')^2 \left( x_{i, p-u} - \sum_{j=1}^{p-w} \frac{\Delta_{n(i)}}{|\Delta_n|} x_{i, j-1} \right) \left( x_{i, p-v} - \sum_{j=1}^{p-w} \frac{\Delta_{n(i)}}{|\Delta_n|} x_{i, j-1} \right).
\]

The adjusting matrices \(A_g\) and \(B_g\) of the gamma working model are derived in the analogous way.