Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Programa de Pós-Graduação em Estatística

Ana Cristina Guedes Pereira

Improved likelihood inference in unit gamma regressions

Recife
2017

Ana Cristina Guedes Pereira

Improved likelihood inference in unit gamma regressions

Master's thesis submitted to the Graduate Program in Statistics, Department of Statistics, Universidade Federal de Pernambuco as a requirement to obtain a Master's degree in Statistics.

Advisor: *Francisco Cribari Neto*
Co-advisor: *Patrícia Leone Espinheira Ospina*

Recife
2017

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

P436i Pereira, Ana Cristina Guedes
Improved likelihood inference in unit gama regressions / Ana Cristina Guedes Pereira. – 2017.
50 f.: il., fig., tab.

Orientador: Francisco Cribari Neto.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2017.
Inclui referências e apêndice.

1. Análise de regressão. 2. Regressão beta. I. Cribari Neto, Francisco (orientador). II. Título.

519.536 CDD (23. ed.) UFPE- MEI 2017-192

ANA CRISTINA GUEDES PEREIRA

IMPROVED LIKELIHOOD INFERENCE IN UNIT GAMMA REGRESSIONS

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 02 de agosto de 2017.

BANCA EXAMINADORA

Prof. Francisco Cribari Neto
UFPE

Prof. Aldo William Medina Garay
UFPE

Prof. Aluísio de Souza Pinheiro
Unicamp

To God and to my mother, Neide.

Acknowledgements

I would like to thank my family, especially my parents, Neide Guedes, Jéssica Guedes and Cecília Guedes, and all my friends for supporting me.
I am very grateful to professors Francisco Cribari and Patrícia Espinheira for their invaluable advices and guidance.
I would also like to thank all professors at Universidade Federal de Pernambuco (UFPE), especially professors Alex Dias, Audrey Cysneiros, Raydonal Ospina and Gauss Cordeiro.
I would like to thank my former professors at Universidade Federal do Ceará (UFC), especially professors Juvêncio Nobre, João Maurício and Ronald Targino, who helped me immensely during my undergraduate studies.
I am thankful to all my colleagues and to the friends I made in Recife, especially César Diogo, Wenia Valdevino, Marcones Sousa, Thiago Oliveira, Alejandro Arias, Yuri Alves, Rodney Fonseca, João Eudes, Vinícius Scher, Jucelino Matos, Bruna Palm, Jessica Barbosa and Alex Santos. I would not like to be unfair with anyone, so I will not list further names, but I would like to thank everyone I spent time with in this great city.
I would like to thank all staff members at Centro de Ciências Exatas e da Natureza (CCEN), especially Valéria for her competent assistance.
Finally, I thank CAPES for the financial support.

Our world, our life, our destiny, are dominated by uncertainty; this is perhaps the only statement we may assert without uncertainty.

—BRUNO DE FINETTI

Abstract

In this dissertation, we focus on the issue of performing likelihood ratio testing inferences in unit gamma regressions. Our interest lies in testing inferences that are accurate and reliable in small samples. The unit gamma regression model was proposed by Mousa et al. (2016) based on the unit gamma distribution introduced by Grassia (1977). Closed form expressions for the score vector and for Fisher's information matrix were obtained by Mousa et al. (2016). The model is useful for dealing with doubly limited continuous dependent variables (DLCDV), such as proportions, indices and rates, being an alternative to the beta regression model, which has been widely used in the literature. We derive a small sample adjustment to the likelihood ration ratio test statistic in the class of unit gamma regressions using the approach proposed by Skovgaard (2001). The numerical evidence we present show that the two corrected tests we propose outperform the standard likelihood ratio test in small samples. A real data example is presented.

Keywords: Beta Regression. Likelihood ratio test. Unit gamma distribution. Unit gamma regression.

Resumo

O foco da presente dissertação reside na realização de testes de hipóteses em regressões gama unitária. O teste da razão de verossimilhanças pode ser consideravelmente impreciso em pequenas amostras. Nosso interesse reside na obtenção de testes que sejam precisos e confiáveis quando o tamanho da amostra é pequeno. A distribuição gama unitária foi proposta por Grassia (1977) e serviu de base para o modelo de regressão gama unitário introduzido por Mousa et al. (2016). O modelo sugerido é útil para modelar variáveis dependentes contínuas duplamente limitadas (VDCDL), como proporções, índices e taxas, sendo uma alternativa ao modelo de regressão beta, que tem sido amplamente utilizado na literatura. Nós derivamos uma correção para a estatística da razão de verossimilhanças nessa classe de modelo utilizando o enfoque desenvolvido por Skovgaard (2001). Com base em tal correção, apresentamos duas estatísticas de teste corrigidas. A evidência numérica que nós apresentamos indica que os testes corrigidos conduzem a inferências mais precisas do que aquelas obtidas com o teste da razão de verossimilhanças padrão em pequenas amostras. Aplicamos os resultados a um conjunto real de dados.

Palavras-chave: Distribuição gama unitária. Regressão beta. Regressão gama unitária. Teste da razão de verossimilhanças.

List of Figures

List of Tables

Contents

CHAPTER 1

Introduction

1.1 Introduction

In several practical situations, whether experimental or observational, there is interest in investigating how a set of variables impacts a given variable of interest which is done through regression analysis. The class of beta regression models proposed by Ferrari and Cribari-Neto (2004) can be used to model dependent variables that assume values in the standard unit interval, i.e., in $(0,1)$, such as rates, proportions, and concentration indices. Many empirical studies in different fields have been based on the beta regression model, such as Brehm and Gates (1993), Hancox et al. (2010), Kieschnick and McCullough (2003), Smithson and Verkuilen (2006) and Zucco (2008). An alternative regression model was proposed by Mousa et al. (2016) that was based on the unit gamma distribution introduced by Grassia (1977). As the beta regression model, the unit gamma regression model allows practitioners to model responses that assume values in the standard unit interval.

The likelihood ratio test is the most commonly used test for making testing inferences in regression analysis. Such a test, however, can be considerably size-distorted in small samples. A well known correction to the likelihood test statistic is the Bartlett correction (Lawley, 1956). A similar correction to the score test statistic was obtained by Cordeiro and Ferrari (1991). Such corrections, however, require the use a large number of log-likelihood cumulants and involve long and tedious algebra. An alternative correction that can be more easily obtained was proposed by Skovgaard (2001). It only requires second-order log-likelihood derivatives. We use such an approach to obtain two modified likelihood ratio test statistics that can be used to perform reliable testing inferences in unit gamma regressions. The Monte Carlo evidence that we present show that testing inferences based on the two modified test statistics can be considerably more accurate than that based on the standard likelihood test statistic in small samples.

1.1.1 The dissertation structure

In Chapter 2, we present the beta regression model and the unit gamma regression model, with their respective score functions and information matrices. In Chapter 3, we derive two modified likelihood test statistics using the approach proposed by Skovgaard (2001). In Chapter 4 we present and discuss the results of a set of Monte Carlo simulations that were performed to evaluate the finite sample behavior of the standard and modified likelihood ratio tests. In Chapter 5 we present an empirical application that uses both the unit gamma and the beta regression model. Some concluding remarks are offered in Chapter 6.

1.1.2 Computational resources

The programming routines for Monte Carlo simulations were written in the OX matrix programming language (Doornik, 2009) version 7.10 for the WINDOWS operating system. OX is an object-oriented matrix programming language whoso syntax is similar to those of C and C++. It is available ree of costs for academic use at `http://www.doornik.com`. For more details, see Doornik (2009). All figures presented in this dissertation were produced using the R statistical computing environment (Team, 2014) version 3.4.0 for the WINDOWS operating system. It is freely available at `http://www.R-project.org`. The dissertation was typeset using LaTeX.

1.2 Maximum likelihood estimation

Consider a random sample $\mathbf{y} = (y_1, \ldots, y_n)$ from a population with probability density function f which is indexed by $\theta \in \Theta \subseteq \mathbb{R}^k$, Θ being the parameter space. The main interest lies in performing inferences on the components of θ, i.e., on $\theta_1, \ldots, \theta_k$. Such inferences are based on statistics, i.e., on functions of the data. Let $\hat{\theta}$ be a statistic that is used to estimate θ, i.e., an estimator of θ. The most commonly used estimation method is the maximum likelihood method. The maximum likelihood estimator, $\hat{\theta}$, is the value of θ that maximizes the likelihood function given by

$$L(\theta) \equiv L(\theta; \mathbf{y}) = f(\mathbf{y}; \theta), \theta \in \Theta, \Theta \subseteq \mathbb{R}^k.$$

Notice that the likelihood function is the joint probability density function, but viewed as a function of θ for a given $\mathbf{y}$. That is,

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L(\theta).$$

It is usually more convenient to maximize the logarithm of the likelihood function, which is known as the log-likelihood function:

$$\ell(\theta) = \log(L(\theta)) = \ell(\theta; \mathbf{y}).$$

Notice that the value of θ that maximizes $L(\theta)$ also maximizes $\ell(\theta)$.

The score function, $U(\theta) = (U_1(\theta), \ldots, U_k(\theta))^\top$, is the vector of log-likelihood derivatives with respect to the unknown parameters, i.e.,

$$U_r(\theta) = \partial \ell(\theta) / \partial \theta_r, \; r = 1, \ldots, k,$$
$$U(\theta) = \bigtriangledown_\theta \{\ell(\theta)\},$$

where $\bigtriangledown_\theta = (\partial/\partial\theta_1, \ldots, \partial/\partial\theta_k)^\top$ is the gradient operator. The score function shows how the log-likelihood function varies with θ.

Maximum likelihood estimators enjoy several desirable properties (Lehmann and Casella, 2011). For instance, they are: (i) asymptotically unbiased, (ii) consistent, (iii) asymptotically normally distributed, and (iv) asymptotically efficient.

1.3 Nonlinear optimization

Many, if not most, maximum likelihood estimators cannot be expressed in closed form, i.e., frequently

$$\frac{\partial \ell(\theta)}{\partial \theta} = 0$$

is a system of nonlinear equations whose solution cannot be expressed in closed form. A commonly used approach is to use a nonlinear optimization method to numerically maximize the log-likelihood function in order to obtain parameter estimates based on a given data set. Commonly used optimization methods are Newton-Raphson, Fisher's score, steepest descent, BHHH and BFGS. The latter belongs to the class of quasi-Newton methods whereas the remaining methods are members of Newton class. Such methods are iterative, i.e., they start at a given point (θ_0) and iterate until convergence is reached. The iterative scheme is of the form

$$\theta_{t+1} = \theta_t + \lambda_t \Delta_t, \quad t = 0, 1, 2, \ldots.$$

Here, λ_t is a positive scalar known as the step length and Δ_t is the directional vector. Ideally, an optimal value for λ_t should be computed at each iteration, i.e., the value that solves

$$\frac{\partial \ell(\theta_t + \lambda_t \Delta_t)}{\partial \lambda_t} = U(\theta_t + \lambda_t \Delta_t)^\top \Delta_t = 0.$$

Since such approach is computational burdensome, it is often replaced by the use of an *ad hoc* rule.

The most used class of methods is the gradient class, for which $\Delta_t = M_t \times U_t$, M_t being a positive-definite matrix and U_t denoting the score function at the t-th iteration. Different choices of M_t yield different optimizations methods.

Let $J \equiv J(\theta)$ denote the Hessian matrix, i.e., the matrix of second log-likelihood derivatives. The Newton-Raphson method uses the following updating scheme:

$$\theta_{t+1} = \theta_t - \lambda_t J_t^{-1} U_t, \quad t = 0, 1, 2, \ldots.$$

A potential problem with this method is that the Hessian is not guaranteed to be negative-definite when we are not close to the maximum which would case M_t not to be positive-definite.

The class of quasi-Newton method overcomes that problem by using an updating scheme for M_t:

$$M_{t+1} = M_t + N_t,$$

where N_t is positive-definite. If M_0 is positive definite, then all elements of the above sequence will also be positive-definite. N_t must be such that M_t converges to $-J_t^{-1}$ as $t \to \infty$. Notice that the methods that belong to the quasi-Newton class do not make use of the Hessian matrix, i.e., they do not require second order log-likelihood derivatives.

Let $\delta_t = \theta_{t+1} - \theta_t$ and $v_t = U_{t+1} - U_t$. The BFGS method is the most commonly used quasi-Newton method. It uses

$$M_{t+1} = M_t + \frac{\delta_t \delta_t'}{\delta_t' v_t} + \frac{M_t v_t v_t' M_t}{v_t' M_t \delta_t} - v_t' M_t v_t \left(\frac{\delta_t}{\delta_t' v_t} - \frac{M_t v_t}{v_t' M_t v_t} \right) \left(\frac{\delta_t}{\delta_t' v_t} - \frac{M_t v_t}{v_t' M_t v_t} \right)'.$$

In the remainder of the dissertation, we shall perform log-likelihood maximizations using the BFGS method. It is implemented into the `MaxBFGS` function of the Ox matrix programming language (Doornik, 2009). For further details on nonlinear optimization methods, see Nocedal and Wright (2006).

CHAPTER 2

The unit gamma regression model

2.1 Introduction

Regression analysis is commonly used to explain the behavior of doubly limited continuous dependent variables (DBCDVs) that assume values in (a,b). It is well known that the linear regression model not is apropriate for modeling such variables since it may yield fitted values that are smaller than a or larger than b. In addition, it fails to account for distributional asymmetry and heteroskedasticity (Ferrari and Cribari-Neto, 2004). There is then need for regression models that are tailored to DBCDVs. The fixed dispersion beta regression model was proposed by Ferrari and Cribari-Neto (2004). A variable dispersion variant of the model was used by Smithson and Verkuilen (2006) and formally introduced by Simas et al. (2010). An alternative model is based on the unit gamma law (Grassia, 1977). Like the beta regression model, the unit gamma regression model has two variants, namely: fixed and variable dispersion variants.

2.1.1 Beta regression

The beta distribution is a parametric alternative to the normal distribution for modeling continuous data that are restricted to the standard unit interval. Its density can be left- or right-skewed, symmetric, J-shaped and inverted J-shaped. Ferrari and Cribari-Neto (2004) proposed a regression model based on the assumption that the response y follows the beta distribution. They changed the distribution parameterization so that it becomes indexed by location (μ) and precision (ϕ) parameters, i.e., $y \sim Beta(\mu,\phi)$. The reparameterized beta density is

$$b(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma[\phi(1-\mu)]} y^{\mu\phi-1}(1-y)^{\phi(1-\mu)-1}, \quad 0<y<1, 0<\mu<1, \phi>0,$$

where $\Gamma(\cdot)$ is the gamma function. Here, $\mathbb{E}(y)=\mu$ and $\text{Var}(y)=\mu(1-\mu)/(\phi+1)$. Some beta densities are displayed in Figure 2.1.

Let $y_1,\ldots,y_n$ be independent random variables, each y_i following the beta law, i.e., $y_i \sim Beta(\mu_i,\phi_i)$, $i=1,\ldots,n$. The location (mean) submodel is

$$g_1(\mu_i) = \eta_i = \sum_{j=1}^{p} \beta_j x_{ij}$$

and the precision submodel is

$$g_2(\phi_i) = \zeta_i = \sum_{j=1}^{q} \delta_j h_{ij},$$

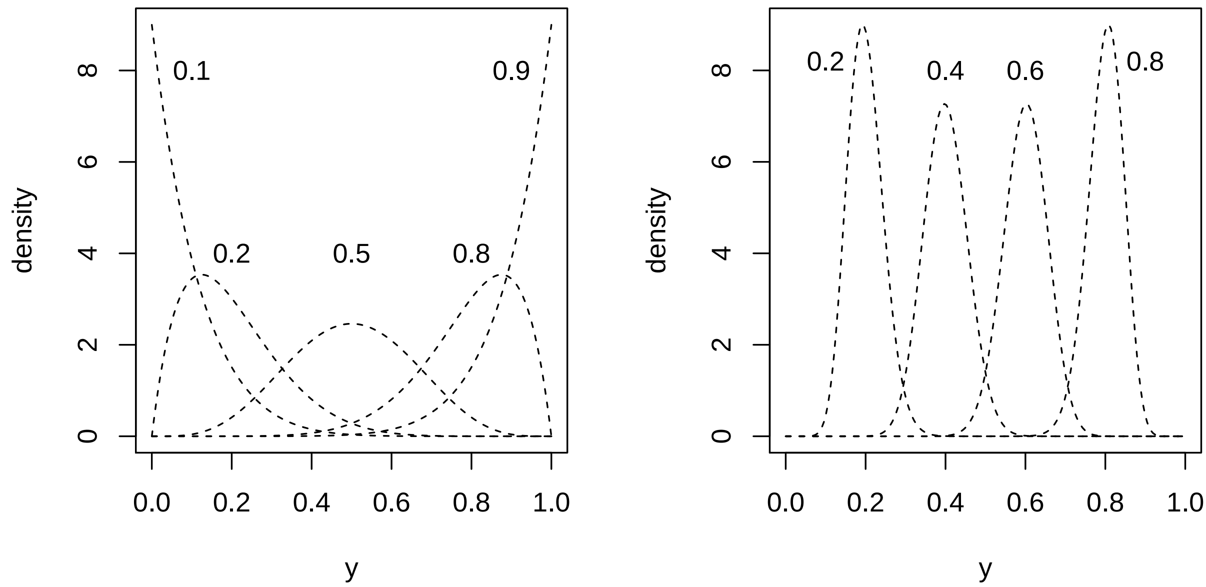

Figure 2.1 Beta densities for different values of μ; $\phi = 10$ (left panel) and $\phi = 80$ (right panel).

where $\beta = (\beta_1, \ldots, \beta_p)^\top$ and $\delta = (\delta_1, \ldots, \delta_q)^\top$ are vectors of unknown regression parameters $\beta \in \mathbb{R}^p$ and $\delta \in \mathbb{R}^q$, $x_{i1} \equiv h_{i1} \equiv 1$, $x_{i2}, \ldots, x_{ip}$, $h_{i2}, \ldots, h_{iq}$ are observations on p and q covariates ($p + q < n$). Finally, $g_1(\cdot)$ and $g_2(\cdot)$ are strictly monotonic and twice differentiable that map $(0,1) \to \mathbb{R}$ and $(0,\infty) \to \mathbb{R}$, respectively. Possible choices of $g_1(\mu_i)$ are: logit $g_1(\mu_i) = \log[\mu_i/(1-\mu_i)]$, probit $g_1(\mu_i) = \Phi^{-1}(\mu_i)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function, log-log $g_1(\mu_i) = -\log[-\log(\mu_i)]$, complementary log-log $g_1(\mu_i) = \log[-\log(1-\mu_i)]$, and Cauchy $g_1(\mu_i) = \tan[\pi(\mu_i - 0.5)]$. Possible choices of $g_2(\phi_i)$ are: identity $g_2(\phi_i) = \phi_i$, log $g_2(\phi_i) = \log(\phi_i)$ and square root $g_2(\phi_i) = \sqrt{\phi_i}$. In the fixed dispersion beta regression model, $g_2(\phi_i) = g_2(\phi) = \delta_0$.

2.2 Unit gamma regression

An alternative regression model was introduced by Mousa et al. (2016). It is based on the unit gamma distribution proposed by Grassia (1977). As the beta regression, it is useful for modeling DBCDVs. The unit gamma distribution is indexed by two parameters: $\alpha, \phi > 0$. We shall denote it by $ug(\alpha, \phi)$. The unit gamma density can be written as

$$ug(y; \alpha, \phi) = \frac{\alpha^\phi}{\Gamma(\phi)} y^{\alpha-1} \log\left(\frac{1}{y}\right)^{\phi-1}, \quad 0 < y < 1, \alpha > 0, \phi > 0.$$

Here, $\mathbb{E}(y) = [\alpha/(\alpha+1)]^{\phi}$ and $\text{Var}(y) = [\alpha/(\alpha+2)]^{\phi} - [\alpha/(\alpha+1)]^{2\phi}$. The density was reparameterized by setting $\alpha = [\mu^{1/\phi}/(1-\mu^{1/\phi})]$ so that $\mathbb{E}(y) = \mu$ and $\text{Var}(y) = \mu\{[1/(2-\mu^{1/\phi})^{\phi}] - \mu\}$. Notice that the variance is a function of the mean. The new density function is

$$ug(y;\mu,\phi) = \frac{\left(\frac{\mu^{1/\phi}}{1-\mu^{1/\phi}}\right)^{\phi}}{\Gamma(\phi)} y^{\frac{\mu^{1/\phi}}{1-\mu^{1/\phi}}-1} \log\left(\frac{1}{y}\right)^{\phi-1}, \quad 0 < y < 1, 0 < \mu < 1, \phi > 0. \qquad (2.1)$$

We shall denote the reparameterized unit gamma distribution by $ug(\mu,\phi)$. Notice that ϕ can be interpreted as a precision parameter. Some unit gamma densities are displayed in Figure 2.2.

Table 2.1 contains values of the variance of y corresponding to different values of μ and ϕ for the beta and unit gamma laws. It is noteworthy that for the same values of the two parameters the variance of y is typically smaller in the unit gamma distribution. For instance, when $\mu = 0.5$ and $\phi = 30.0$, we obtain 0.0081 for the beta law and and 0.0039 for the unit gamma law.

Table 2.1 Variance of y for different values of μ and ϕ.

	beta distribution					unit gamma distribution				
ϕ \ μ	0.1	0.3	0.5	0.8	0.9	0.1	0.3	0.5	0.8	0.9
0.5	0.0600	0.1400	0.1667	0.1067	0.0600	0.0609	0.1271	0.1280	0.0460	0.0150
1.0	0.0450	0.1050	0.1250	0.0800	0.0450	0.0426	0.0865	0.0833	0.0267	0.0082
2.0	0.0300	0.0700	0.0833	0.0533	0.0300	0.0253	0.0522	0.0491	0.0145	0.0043
5.0	0.0150	0.0350	0.0417	0.0267	0.0150	0.0108	0.0238	0.0220	0.0061	0.0018
10.0	0.0082	0.0191	0.0227	0.0145	0.0082	0.0054	0.0124	0.0115	0.0031	0.0009
15.0	0.0056	0.0131	0.0156	0.0100	0.0056	0.0036	0.0084	0.0078	0.0021	0.0006
30.0	0.0029	0.0068	0.0081	0.0052	0.0029	0.0018	0.0043	0.0039	0.0011	0.0003
50.0	0.0018	0.0041	0.0049	0.0031	0.0018	0.0011	0.0026	0.0024	0.0006	0.0002
70.0	0.0013	0.0030	0.0035	0.0023	0.0013	0.0008	0.0019	0.0017	0.0005	0.0001
90.0	0.0010	0.0023	0.0027	0.0018	0.0010	0.0006	0.0014	0.0013	0.0004	0.0001

Let $y_1,\ldots,y_n$ be independent random variables, where each $y_i \sim ug(\mu_i,\phi_i)$, $i = 1,\ldots,n$, with mean μ_i and precision ϕ_i. In the unit gamma regression model proposed by Mousa et al. (2016), the ith mean response can be written as

$$g_1(\mu_i) = \eta_i = \sum_{j=1}^{p} \beta_j x_{ij}, \qquad (2.2)$$

the ith precision being given by

$$g_2(\phi_i) = \zeta_i = \sum_{j=1}^{q} \delta_j h_{ij}. \qquad (2.3)$$

Since $\text{Var}(y_i) = \mu_i\{[1/(2-\mu_i^{1/\phi_i})^{\phi_i}] - \mu_i\}$, where $\mu_i = g_1^{-1}(\eta_i)$, the regression model is heteroskedastic. The fixed dispersion unit gamma regression model is obtained by setting $g_2(\phi_i) = g_2(\phi) = \delta_0$.

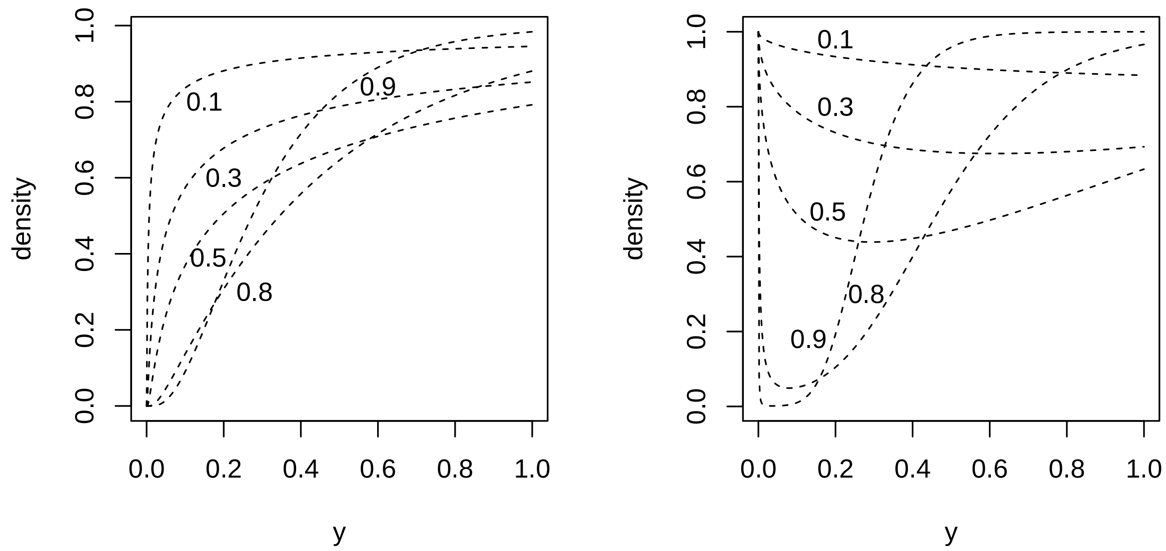

Figure 2.2 Unit gamma densities for different values of μ; $\phi = 0.5$ (left panel) and $\phi = 1.5$ (right panel).

Consider the unit gamma regression model defined in (2.2) and (2.3) and write the model parameter vector as $\theta = (\beta^\top, \delta^\top)^\top$. The variable dispersion unit gamma regression model log-likelihood function is

$$\ell(\theta) \equiv \ell(\beta, \delta) = \sum_{i=1}^{n} \ell_i(\mu_i, \phi_i), \tag{2.4}$$

where

$$\ell_i(\mu_i, \phi_i) = \phi_i \log(d_i) - \log\Gamma(\phi) + (d_i - 1)y_i^* + (\phi_i - 1)y_i^\dagger.$$

Here

$$y_i^* = \log(y_i), \quad y_i^\dagger = \log(-\log(y_i)) \quad \text{and} \quad d_i = \mu_i^{1/\phi_i}/(1-\mu_i^{1/\phi_i}). \tag{2.5}$$

Let $z = -y^* = -\log(y)$. The distribution function of z can be obtained as follows:

$$F_z(a) = \Pr(z \leq a) = \Pr(-\log(y) \leq a) = \Pr(y > \exp(-a)) = 1 - F_y(\exp(-a)).$$

The above result implies that

$$f_z(a) = -f_y(\exp(-a)) \times (-\exp(-a)) = \frac{d^\phi}{\Gamma(\phi)} a^{(\phi-1)} \exp(-da).$$

Hence, $z = -y^*$ is gamma-distributed with parameters ϕ and d. It then follows that $\mathbb{E}(y^*) = \mathbb{E}(-z) = -\phi/d$. Additionally, $y^\dagger = \log(-\log y) = \log(z)$ is distributed as log-gamma with parameters ϕ and d. It can then be shown that

$$\mu_i^* = \mathbb{E}(y_i^*) = -\frac{\phi_i}{d_i}, \quad \mu_i^\dagger = \mathbb{E}(y_i^\dagger) = \psi(\phi_i) - \log(d_i),$$
$$\upsilon_i^* = \mathrm{Var}(y_i^*) = \frac{\phi_i}{d_i^2}, \quad \upsilon_i^\dagger = \mathrm{Var}(y_i^\dagger) = \psi'(\phi_i), \quad c^{*\dagger} = -\frac{1}{d_i}, \tag{2.6}$$

where $\psi(\cdot)$ and $\psi'(\cdot)$ are the digamma and trigamma functions, respectively.

The log-likelihood function $\ell(\theta)$ can be written in matrix form as

$$\ell(\theta) = \{(y^* - \mu^*)^\top (D - \mathscr{I}) + (y^\dagger - \mu^\dagger)^\top (\Phi - \mathscr{I}) + b^\top\}\iota, \tag{2.7}$$

where $y^* = (y_1^*, \ldots, y_n^*)^\top$, $\mu^* = (\mu_1^*, \ldots, \mu_n^*)^\top$, $y^\dagger = (y_1^\dagger, \ldots, y_n^\dagger)^\top$, $\mu^\dagger = (\mu_1^\dagger, \ldots, \mu_n^\dagger)^\top$, $D = \mathrm{diag}(d_1, \ldots, d_n)$, $\Phi = \mathrm{diag}(\phi_1, \ldots, \phi_n)$, $\mathscr{I}$ is the $n \times n$ identity matrix, ι is the n-dimensional column vector of ones and $b = (b_1, \ldots, b_n)^\top$ with $b_i = \phi_i \log(d_i) - \log\Gamma(\phi_i) + \mu_i^*(d_i - 1) + \mu_i^\dagger(\phi_i - 1)$, $i = 1, \ldots, n$, with d_i defined in (2.5).

The first and second derivatives of (2.7) are

(i)

$$\frac{\partial \ell_i(\mu_i, \phi_i)}{\partial \mu_i} = \frac{d_i(1 + d_i)(y_i^* - \mu_i^*)}{\mu_i \phi_i}.$$

(ii)

$$\frac{\partial^2 \ell_i}{\partial \mu_i^2} = \frac{d_i(1 + d_i)}{\phi_i \mu_i^2}\left(\frac{1 + 2d_i}{\phi_i} - 1\right)(y_i^* - \mu_i^*) - \frac{d_i(1 + d_i)}{\mu^{2+1/\phi}\phi}.$$

(iii)

$$\frac{\partial \ell_i(\mu_i, \phi_i)}{\partial \phi_i} = \left(\frac{-d_i \log(\mu_i)(1 + d_i)}{\phi_i^2}\right)(y_i^* - \mu_i^*) + (y_i^\dagger - \mu_i^\dagger).$$

(iv)

$$\frac{\partial^2 \ell_i}{\partial \phi_i^2} = \left[\frac{d_i}{\phi_i^3}\left(\frac{\log(\mu_i)}{\phi_i} + \frac{d_i \log(\mu_i)}{\phi_i} + 2\right)\log(\mu_i)(1 + d_i) + \frac{d_i^2(\log(\mu_i))^2}{\phi_i^4}(1 + d_i)\right](y_i^* - \mu_i^*)$$
$$- \frac{d_i}{\phi_i^2}\log(\mu_i)(1 + d_i)\left(\frac{1}{d_i} + \frac{\log(\mu_i)}{\phi_i} + \frac{\log(\mu_i)}{d_i \phi_i}\right) - \psi'(\phi_i) - \frac{\log(\mu_i)(1 + d_i)}{\phi_i^2}.$$

(v)

$$\frac{\partial^2 \ell_i}{\partial \mu_i \partial \phi_i} = \left[\frac{-d_i(1 + d_i)\log(\mu_i)}{\mu_i \phi_i}\left(\frac{1 + 2d_i}{\phi_i^2}\right) - \frac{d_i(1 + d_i)}{\mu_i \phi_i^2}\right](y_i^* - \mu_i^*)$$
$$+ \frac{d_i(1 + d_i)}{\mu_i \phi_i}\left(\frac{1}{d_i} + \frac{\log(\mu_i)}{\phi_i} + \frac{\log(\mu_i)}{\phi_i d_i}\right).$$

Under the regularity conditions listed in Sen et al. (2010), the expected value of $\partial\ell_i(\mu_i,\phi_i)/\partial\mu_i$ equals zero. Hence,

$$\mathbb{E}\left[\frac{\partial\ell_i(\mu_i,\phi_i)}{\partial\mu_i}\right]=0\Longleftrightarrow\mathbb{E}\left[\frac{d_i(1+d_i)(y_i^*-\mu_i^*)}{\mu_i\phi_i}\right]=0\Longleftrightarrow\mathbb{E}[y_i^*]=\mu_i^*$$

Using (2.2) and (2.3), we obtain

$$\frac{\partial^2\mu_i}{\partial\eta_i\partial\mu_i}=-\frac{g_1''(\mu_i)}{g_1'(\mu_i)^2}\text{ and }\frac{\partial^2\phi_i}{\partial\zeta_i\partial\phi_i}=-\frac{g_2''(\phi_i)}{g_2'(\phi_i)^2}.$$

It follows from (2.2) and (2.7), $j=1,\ldots,p$, that

$$\frac{\partial\ell(\beta,\phi)}{\partial\beta_j}=\sum_{i=1}^n\left\{\frac{\partial\ell_i(\mu_i,\phi_i)}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\beta_j}\right\}=\sum_{i=1}^n\left\{\frac{d_i}{\mu_i\phi_i}(1+d_i)(y_i^*-\mu_i^*)\frac{1}{g_1'(\mu_i)}x_{ij}\right\}.\tag{2.8}$$

In similar fashion,

$$\begin{aligned}\frac{\partial\ell(\beta,\phi)}{\partial\delta_j}&=\sum_{i=1}^n\left\{\frac{\partial\ell_i(\mu_i,\phi_i)}{\partial\phi_j}\frac{\partial\phi_j}{\partial\zeta_j}\frac{\partial\zeta_j}{\partial\delta_j}\right\}\\&=\sum_{i=1}^n\left\{\left[\left(\frac{-d_i\log(\mu_i)(1+d_i)}{\phi_i^2}\right)(y_i^*-\mu_i^*)+(y_i^\dagger-\mu_i^\dagger)\right]\frac{1}{g_2'(\phi_i)}h_{ij}\right\},\end{aligned}\tag{2.9}$$

$j=1,\ldots,p$.

In what follows we shall write (2.8) and (2.9) in matrix form. The score function, obtained by differentiating the log-likelihood function with respect to the unknown parameters, is given by $U=(U_\beta(\beta,\delta)^\top,U_\delta(\beta,\delta)^\top)^\top$, where

$$U_\beta(\beta,\delta)=X^\top T_1\Phi^{-1}\mathscr{M}^{-1}D(\mathscr{I}+D)(y^*-\mu^*)\tag{2.10}$$

and

$$U_\delta(\beta,\delta)=H^\top T_2[P(y^*-\mu^*)+(y^\dagger-\mu^\dagger)],\tag{2.11}$$

where $T_1=\mathrm{diag}\,(1/g_1'(\mu_1),\ldots,1/g_1'(\mu_n))$, $T_2=\mathrm{diag}\,(1/g_2'(\phi_1),\ldots,1/g_2'(\phi_n))$, $\mathscr{M}=\mathrm{diag}(\mu_1,\ldots,\mu_n)$, $D=\mathrm{diag}(d_1,\ldots,d_n)$, $P=\mathrm{diag}(-d_1(1+d_1)\log(\mu_1)/\phi_1^2,\ldots,-d_n(1+d_n)\log(\mu_n)/\phi_n^2)$, X is an $n\times p$ matrix whose ith row is $x_i^\top$ and H is $n\times q$ matrix whose ith row is $h_i^\top$. The maximum likelihood estimators of β and δ solve $U_\beta(\beta,\delta)=U_\delta(\beta,\delta)=0$. Maximum likelihood estimates are typically obtained by numerically maximizing the log-likelihood function using a Newton or quasi-Newton algorithm such as Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm; see Nocedal and Wright (2006).

We shall now obtain Fisher's information matrix for (β,δ). For $t,j=1,\ldots,p$,

$$\begin{aligned}\frac{\partial^2\ell(\beta,\phi)}{\partial\beta_t\partial\beta_j}&=\sum_{i=1}^n\left\{\left[\frac{\partial^2\ell_i(\mu_i,\phi_i)}{\partial\mu_i^2}\frac{\partial\mu_i}{\partial\eta_i}+\frac{\partial\ell_i(\mu_i,\phi_i)}{\partial\mu_i}\frac{\partial^2\mu_i}{\partial\eta_i\partial\mu_i}\right]\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\beta_j}x_{it}\right\}\\&=\sum_{i=1}^n\left\{\left[\left(\frac{d_i(1+d_i)}{\phi_i\mu_i^2}\left(\frac{1+2d_i}{\phi_i}-1\right)(y_i^*-\mu_i^*)-\frac{d_i(1+d_i)}{\mu^{2+1/\phi}\phi}\right)\frac{1}{g_1'(\mu_i)}\right.\right.\end{aligned}$$

$$+\frac{d_i}{\mu_i\phi_i}(1+d_i)(y_i^*-\mu_i^*)\left(-\frac{g_1''(\mu_i)}{g_1'(\mu_i)^2}\right)\bigg]\frac{1}{g_1'(\mu_i)}x_{ij}x_{it}\bigg\}.$$

Also, for $t = 1,\ldots,p$, $l = 1,\ldots,q$,

$$\begin{aligned}\frac{\partial^2\ell(\beta,\phi)}{\partial\beta_t\partial\delta_l} &= \sum_{i=1}^n\left\{\left[\frac{\partial^2\ell_i(\mu_i,\phi_i)}{\partial\mu_i\partial\phi_i}\frac{\partial\phi_i}{\partial\zeta_i}\frac{\partial\zeta_i}{\partial\delta_l}\right]\frac{1}{g_1'(\mu_i)}x_{it}\right\}\\ &= \sum_{i=1}^n\Bigg\{\Bigg[\left(\frac{-d_i(1+d_i)\log(\mu_i)}{\mu_i\phi_i}\left(\frac{1+2d_i}{\phi_i^2}\right)-\frac{d_i(1+d_i)}{\mu_i\phi_i^2}\right)(y_i^*-\mu_i^*)\\ &\quad+\frac{d_i(1+d_i)}{\mu_i\phi_i}\left(\frac{1}{d_i}+\frac{\log(\mu_i)}{\phi_i}+\frac{\log(\mu_i)}{\phi_i d_i}\right)\Bigg]\frac{1}{g_1'(\mu_i)}\frac{1}{g_2'(\phi_i)}x_{it}h_{it}\Bigg\}.\end{aligned}$$

Finally, for δ_l e $\delta_{l'}$, $l,l' = 1,\ldots,q$,

$$\begin{aligned}\frac{\partial^2\ell(\beta,\phi)}{\partial\delta_l\partial\delta_{l'}} &= \sum_{i=1}^n\left\{\left[\frac{\partial^2\ell_i(\mu_i\phi_i)}{\partial\phi^2}\frac{\partial\phi_i}{\partial\zeta_i}+\frac{\partial\ell_i(\mu_i,\phi_i)}{\partial\phi_i}\frac{\partial^2\phi_i}{\partial\zeta_i\partial\phi_i}\right]\frac{1}{g_2'(\phi_i)}h_{il}h_{il'}\right\}\\ &= \sum_{i=1}^n\Bigg\{\Bigg[\Bigg[\left(\frac{d_i}{\phi_i^3}\left(\frac{\log(\mu_i)}{\phi_i}+\frac{d_i\log(\mu_i)}{\phi_i}+2\right)\log(\mu_i)(1+d_i)+\frac{d_i^2(\log(\mu_i))^2}{\phi_i^4}(1+d_i)\right)\\ &\times(y_i^*-\mu_i^*)-\frac{d_i}{\phi_i^2}\log(\mu_i)(1+d_i)\left(\frac{1}{d_i}+\frac{\log(\mu_i)}{\phi_i}+\frac{\log(\mu_i)}{d_i\phi_i}\right)-\psi'(\phi_i)\\ &-\frac{\log(\mu_i)(1+d_i)}{\phi_i^2}\Bigg]\frac{1}{g_2'(\phi_i)}+\left[\left(\frac{-d_i\log(\mu_i)(1+d_i)}{\phi_i^2}\right)(y_i^*-\mu_i^*)+(y_i^\dagger-\mu_i^\dagger)\right]\\ &\times\left(-\frac{g_2''(\phi_i)}{g_2'(\phi_i)^2}\right)\Bigg]\frac{1}{g_2'(\phi_i)}h_{il}h_{il'}\Bigg\}.\end{aligned}$$

In order to simplify the notation, we define the following matrices:

$$K_1 = \mathrm{diag}\left(\frac{d_i(1+d_i)}{\phi_i\mu_i^2}\left(\frac{1+2d_i}{\phi_i}-1\right)\right),$$

$$K_2 = \mathrm{diag}\left(\frac{d_i(1+d_i)}{\mu^{2+1/\phi}\phi}\right),\quad K_3 = \mathrm{diag}\left(\frac{d_i}{\mu_i\phi_i}(1+d_i)\right),$$

$$Z_1 = \mathrm{diag}\left(\frac{-d_i(1+d_i)\log(\mu_i)}{\mu_i\phi_i}\left(\frac{1+2d_i}{\phi_i^2}\right)-\frac{d_i(1+d_i)}{\mu_i\phi_i^2}\right),$$

$$Z_2 = \mathrm{diag}\left(\frac{d_i(1+d_i)}{\mu_i\phi_i}\left(\frac{1}{d_i}+\frac{\log(\mu_i)}{\phi_i}+\frac{\log(\mu_i)}{\phi_i d_i}\right)\right),$$

$$L_1 = \mathrm{diag}\left(\frac{d_i}{\phi_i^3}\left(\frac{\log(\mu_i)}{\phi_i}+\frac{d_i\log(\mu_i)}{\phi_i}+2\right)\log(\mu_i)(1+d_i)+\frac{d_i^2(\log(\mu_i))^2}{\phi_i^4}(1+d_i)\right),$$

$$L_2 = \text{diag}\left(\frac{d_i}{\phi_i^2}\log(\mu_i)(1+d_i)\left(\frac{1}{d_i}+\frac{\log(\mu_i)}{\phi_i}+\frac{\log(\mu_i)}{d_i\phi_i}\right)-\psi'(\phi_i)-\frac{\log(\mu_i)(1+d_i)}{\phi_i^2}\right),$$

and

$$L_3 = \text{diag}\left(\frac{-d_i\log(\mu_i)(1+d_i)}{\phi_i^2}\right),$$

$i=1,\ldots,n$.

The Hessian matrix, i.e., the matrix of second derivatives is

$$J = \begin{bmatrix} J_{\beta\beta} & J_{\beta\delta} \\ J_{\delta\beta} & J_{\delta\delta} \end{bmatrix},$$

where

$$J_{\beta\beta} = \frac{\partial^2\ell(\beta,\delta)}{\partial\beta\partial\beta^\top} = X^\top[(K_1(Y^*-M^*)-K_2)T_1 - S_1T_1^2K_3(Y^*-M^*)]T_1X,$$

$$J_{\beta\delta} = \frac{\partial^2\ell(\beta,\delta)}{\partial\beta\partial\delta^\top} = J_{\delta\beta}^\top = X^\top[Z_1(Y^*-M^*)+Z_2]T_1T_2H,$$

and

$$J_{\delta\delta} = \frac{\partial^2\ell(\beta,\delta)}{\partial\delta\partial\delta^\top} = H^\top[T_2(L_1(Y^*-M^*)+L_2)-(L_3(Y^*-M^*)+(Y^\dagger-M^\dagger))S_2T_2^2]T_2H.$$

Here, $Y^*=\text{diag}(y_1^*,\ldots,y_n^*)$, $Y^\dagger=\text{diag}(y_1^\dagger,\ldots,y_n^\dagger)$, $M^*=\text{diag}(\mu_1^*,\ldots,\mu_n^*)$, $M^\dagger=\text{diag}(\mu_1^\dagger,\ldots,\mu_n^\dagger)$, $S_1=\text{diag}\left(g_1''(\mu_1),\ldots,g_1''(\mu_n)\right)$ and $S_2=\text{diag}\left(g_2''(\phi_i),\ldots,g_2''(\phi_i)\right)$.

We can write (2.1) as

$$ug(y;\alpha,\phi) = \exp\left(\phi\log\left(\frac{\mu^{1/\phi}}{1-\mu^{1/\phi}}\right)-\log\Gamma(\phi)-\left(\frac{\mu^{1/\phi}}{1-\mu^{1/\phi}}\right)y^* + \phi y^\dagger\right)\frac{1}{-y\log(y)}. \tag{2.12}$$

Such a density can be also written as

$$p(y|\vartheta) = \exp\left(\sum_{i=1}^{2}\vartheta_iT_i(y)-A(\vartheta)\right)h(y),$$

where $(\vartheta_1,\vartheta_2)=\left(\frac{\mu^{1/\phi}}{1-\mu^{1/\phi}},\phi\right)$, $(T_1(y),T_2(y))=(y^*,y^\dagger)$ and

$$A(\vartheta_1,\vartheta_2) = -\phi\log\left(\frac{\mu^{1/\phi}}{1-\mu^{1/\phi}}\right)+\log\Gamma(\phi).$$

It is thus clear that the unit gamma density belongs to the two-dimensional exponential family. As a consequence, the standard regularity conditions (e.g. 6.3, Lehmann and Casella (2011)) hold and it follows that Fisher's information matrix equals $\mathbb{E}(-J)$, where $-J$ is the observed information matrix. Fisher's information matrix is thus given by

$$I = I(\beta,\delta) = \begin{bmatrix} I_{\beta\beta} & I_{\beta\delta} \\ I_{\delta\beta} & I_{\delta\delta} \end{bmatrix},$$

where

$$I_{\beta\beta} = X^\top K_2 T_1^2 X, I_{\beta\delta} = I_{\delta\beta}^\top = -X^\top Z_2 T_1 T_2 H \text{ and } I_{\delta\delta} = -H^\top L_2 T_2^2 H.$$

It is noteworthy that, unlike what happens in the class of generalized linear models (McCullagh and Nelder, 1989), the parameters β and δ are not orthogonal. We also note that as shown by Lehmann and Casella (2011)), for densities that can be written as in (2.12),

$$\mathbb{E}_\vartheta(T_j) = \frac{\partial A(\vartheta)}{\partial \vartheta_j} \text{ and } \mathrm{Cov}_\vartheta(T_j,T_k) = \frac{\partial^2 A(\vartheta)}{\partial \vartheta_j \partial \vartheta_k}$$

which provides an alternative way for obtaining the quantities in (2.6).

Under standard regularity conditions and when the sample size is large,

$$\begin{pmatrix} \hat{\beta} \\ \hat{\phi} \end{pmatrix} \sim N_{p+1}\left(\begin{pmatrix} \beta \\ \phi \end{pmatrix}, I^{-1}\right).$$

It is then possible to obtain standard errors for the maximum likelihood estimates by computing the square roots of the diagonal elements of I^{-1} after replacing the unknown parameters by the corresponding estimates. The asymptotic normality of $(\hat{\beta}^\top, \hat{\phi})$ is also useful for interval estimation and hypothesis testing inference. In the following chapters we shall focus on hypothesis testing inference on the parameters that index the unit gamma regression model.

CHAPTER 3

Modified likelihood ratio testing inference in unit gamma regressions

Our interest lies in performing testing inferences on the parameter that index a given model. Oftentimes no exact test is available, and it is necessary to use a test based on an approximation, more especifically, a test whose null distribution of the associated test statistic is approximated by a parameter-free distribution. Commonly used tests are the likelihood ratio, score and Wald tests. When the sample size is small such tests tend to be size distorted, i.e., they typically display poor control of the type I error frequency. Different approachs have been proposed in the literature to circumvent such a shortcoming.

Ferrari and Cribari-Neto (2004) discussed likelihood ratio, score and Wald testing inference in the class of beta regression models. In order to achieve more accurate inferences in small samples, Ferrari and Pinheiro (2011) derived the adjustment introduced by Skovgaard (2001) to the likelihood ratio test statistic. A similar adjustment was derived by Pereira and Cribari-Neto (2014) for the class of inflated beta regressions. The authors obtained adjustments to the likelihood ratio and signed likelihood ratio test statistics and showed that the modified tests are typically considerably more accurate than the corresponding unmodified tests.

In this chapter we shall present the likelihood ratio test and obtain two corrected likelihood ratio test statistics in the class of unit gamma regression models. The modified tests are expected to be more accurate than the standard likelihood ratio test in small samples.

3.1 The likelihood ratio test

Consider the unit gamma regression model in Equations (2.2)–(2.3) and the corresponding log-likelihood function presented in (2.4), where $\theta = (\beta^\top, \delta^\top)^\top$ is the model k-dimensional parameter vector, β being a p-vector and δ being a q-vector such that $p+q=k$.

In what follows, $\kappa = (\kappa_1, \ldots, \kappa_l)^\top$ represents the parameter of interest and $\psi = (\psi_1, \ldots, \psi_s)^\top$ is the nuisance parameter. (Note that $l+s=p+q$). Our interest, thus, shall lie in testing l restrictions. In particular, we wish to test $\mathscr{H}_0 : \kappa = \kappa^0$ versus $\mathscr{H}_0 : \kappa \neq \kappa^0$, where κ^0 is a fixed l-vector. The likelihood ratio test statistic can be written as

$$w = 2[\ell(\hat{\kappa}, \hat{\psi}) - \ell(\kappa^0, \tilde{\psi})],$$

where $(\kappa^{0\top}, \tilde{\psi}^\top)$ and $(\hat{\kappa}^\top, \hat{\psi}^\top)$ are, respectively, the restricted and unrestricted maximum likelihood estimators of $(\kappa^\top, \psi^\top)$. Under the null hypothesis, w is asymptotically distributed as χ^2_l. The null hypothesis is rejected at the α significance level ($0 < \alpha < 1$) if $w > \chi^2_{1-\alpha,l}$, where $\chi^2_{1-\alpha,l}$ is the $1-\alpha$ upper χ^2_l quantile.

When n is small, the approximation used in the likelihood ratio test may not be accurate, and as a result size distortions may take place. It is thus desirable to apply a finite sample correction to the test statistic when performing inferences with samples of small sizes. As noted earlier, a useful small sample correction was obtained by Skovgaard (2001). Our chief goal is to obtain adjusted likelihood ratio test statistics whose distributions under $\mathscr{H}_0$ are well approximated by the χ^2 reference distribution even in small samples.

3.2 Adjusted test statistics

As explained in the previous section, likelihood ratio testing inference relies on an asymptotic approximation: the test statistic null distribution is approximated by its limiting counterpart, the χ^2_l distribution, where l is the number of restrictions under test. The critical value used in the test is obtained from the test statistic asymptotic null distribution. In small samples, such an approximation may be poor and, as a result, considerably size distortions may take place. Several strategies have been developed in the literature to overcome such a shortcoming. One of such approaches involve multiplying the test statistic by a Bartlett correction factor (Lawley, 1956) which typically results in more accurate inferences. The derivation of such a correction factor is oftentimes, however, quite cumbersome. For details on the Bartlett correction, see Cribari-Neto and Cordeiro (1996) and the references therein.

When the interest lies in making inferences on a subset of the parameter vector, i.e., when there are nuisance parameteres, one can use marginal or conditional likelihood functions. In some models, it is possible to obtain a profile likelihood function, but such a function does not enjoy the same properties as the usual likelihood ratio function. Several corrections were proposed in the literature aiming at reducing the impact of the nuisance parameters on the inference made on the parameters of interest; see, e.g., Barndorff-Nielsen (1983), Barndorff-Nielsen (1994), Cox and Reid (1987),Cox and Reid (1992), McCullagh and Tibshirani (1990) and Stern (1997). Such corrections tend reduce the curvature of the log-likelihood of the function.

When the parameter of interest is scalar, the null hypothesis can be tested using the signed likelihood ratio test statistic:

$$w_s = \mathrm{sinal}(\hat{\kappa} - \kappa)[2\{\ell(\hat{\kappa}, \hat{\psi}) - \ell(\kappa, \tilde{\psi})\}]^{1/2},$$

where κ is the l-vector of parameters of interest and ψ is the s-vector of nuisance parameters. Under the null hypothesis, w_s is standard normally distributed with error of order $n^{-1/2}$. Such an approximation may be, nonetheless, inaccurate in small sample sizes.

Barndorff-Nielsen (1986) and Barndorff-Nielsen (1991) introduced the following modified signed likelihood ratio test statistic:

$$w_s^* = w_s + w_s^{-1} \log \varepsilon, \tag{3.1}$$

where

$$\varepsilon = |\hat{J}|^{1/2} |\tilde{U}'|^{-1} |\tilde{J}_{\psi\psi}|^{1/2} \frac{w_s}{[(\hat{\ell}' - \tilde{\ell}')^\top (\tilde{U}')^{-1}]_\kappa}, \tag{3.2}$$

U being the score function and J being the matrix of second-order log-likelihood derivatives. Notice that $-J$ is the observed information matrix. Here, $\tilde{U}'$ and $\tilde{\ell}'$ are, respectively, the score vector and log-likelihood derivatives with respect to $\hat{\theta}$, both evaluated at $\tilde{\theta}$. Similarly, $\hat{\ell}'$ denotes the log-likelihood derivative with respect to $\hat{\theta}$ evaluated at $\hat{\theta}$. Additionally, $J_{\psi\psi}$ denotes the $s \times s$ matrix of second derivatives with respect to ψ and $[(\hat{\ell}' - \tilde{\ell}')^\top (\tilde{U}')^{-1}]_\kappa$ is the element of vector $(\hat{\ell}' - \tilde{\ell}')^\top (\tilde{U}')^{-1}$ corresponding to the parameter κ, which in this case is scalar. It is noteworthy that the derivation of such quantities is oftentimes quite cumbersome. The proposed test statistic is standard normally distributed under the null hypothesis with error of order $n^{-3/2}$.

Skovgaard (1996) obtained approximations to the sample space derivatives required to compute the test statistic given in (3.1) for when the parameter of interest is scalar. His results were generalized for vector-valued parameter of interest in Skovgaard (2001). Several authors have recently obtained such an adjustment for different classes of models; see, e.g., Ferrari and Cysneiros (2008), Melo et al. (2009) and Pereira (2010). In particular, we note that the Skovgaard adjustment was derived by Ferrari and Pinheiro (2011) for performing testing inferences in the beta regression model. In what follows, we shall derive a similar adjustment for likelihood ratio testing inference in the unit gamma regression model.

By approximating $\tilde{U}'$, $\tilde{\ell}'$ and $\hat{\ell}'$, Skovgaard (2001) obtained the following adjusted likelihood ratio statistic:

$$w^* = w - 2\log\xi, \tag{3.3}$$

where

$$\xi = \frac{\{|\tilde{I}||\hat{I}||\tilde{J}_{\psi\psi}|\}^{1/2}}{|\bar{\Upsilon}||\{\tilde{I}\bar{\Upsilon}^{-1}\hat{J}\hat{I}^{-1}\bar{\Upsilon}\}_{\psi\psi}|^{1/2}} \frac{\{\tilde{U}^\top\bar{\Upsilon}^{-1}\hat{I}\hat{J}^{-1}\bar{\Upsilon}\tilde{I}^{-1}\tilde{U}\}^{l/2}}{w^{l/2-1}\tilde{U}^\top\bar{\Upsilon}^{-1}\bar{q}}, \tag{3.4}$$

$J_{\psi\psi}$ being the $s \times s$ matrix of second derivatives with respect to ψ. As noted earlier, when the relevant regularity conditions are satisfied, the negative Hessian, $-J_{\psi\psi}$, is the observed information matrix relative to ψ. Note that $\bar{q}$ is a vector of dimension $l+s$ and $\bar{\Upsilon}$ is a matrix of dimension $(l+s) \times (l+s)$. Under $\mathscr{H}_0$, w^* is asymptotically distributed as χ^2_l. The quantities $\bar{q}$ and $\bar{\Upsilon}$ come from

$$q = \mathrm{E}[U(\theta_1)(\ell(\theta_1) - \ell(\theta))]$$

and

$$\Upsilon = \mathrm{E}[U(\theta_1)U^\top(\theta)],$$

by replacing θ_1 with $\hat{\theta}$ and θ with $\tilde{\theta}$ after the expected values are computed. Here, $\hat{\theta}$ and $\tilde{\theta}$ denote, respectively, the unrestricted and restricted maximum likelihood estimators of θ.

A test statistic which is asymptotically equivalent to w^* is

$$w^{**} = w\left(1 - \frac{1}{w}\log\xi\right)^2. \tag{3.5}$$

A clear advantage of w^{**} is that it is always non-negative. The numerical evidence in Skovgaard (2001) shows that w^* may slightly outperform w^{**} in some cases. Under $\mathscr{H}_0$ both test statistics are χ^2_l distributed with high degree of accuracy. The adjusted statistics are invariant under reparametrizations of the form $(\kappa,\psi) \longrightarrow (\kappa,\varphi(\kappa,\psi))$. For further details, see Skovgaard (2001).

In what follows, we shall obtain closed form expressions for the quantities $\bar{q}$ and $\bar{\Upsilon}$ in the class of unit gamma regression models. Note that $\bar{q}$ is obtained from

$$q = \begin{bmatrix} \mathbb{E}[U_\beta(\theta_1)\ell(\theta_1)] - \mathbb{E}[U_\beta(\theta_1)\ell(\theta)] \\ \mathbb{E}[U_\delta(\theta_1)\ell(\theta_1)] - \mathbb{E}[U_\delta(\theta_1)\ell(\theta)] \end{bmatrix},$$

where $\theta = (\beta^\top, \delta^\top)^\top$.

By using Equations (2.7) and (2.10) we obtain

$$\begin{aligned}
\mathbb{E}[U_\beta(\theta)\ell(\theta)] &= \mathbb{E}_\theta[X^\top T_1 \Phi^{-1}\mathscr{M}^{-1} D(\mathscr{I}+D)(y^* - \mu^*) \\
&\quad \times \{(y^*-\mu^*)^\top (D-\mathscr{I}) + (y^\dagger - \mu^\dagger)^\top(\Phi - \mathscr{I}) + b^\top\}\iota] \\
&= X^\top T_1 \Phi^{-1}\mathscr{M}^{-1} D(\mathscr{I}+D)\{\mathbb{E}_\theta[(y^*-\mu^*)(y^*-\mu^*)^\top](D-\mathscr{I}) \\
&\quad + \mathbb{E}_\theta[(y^\dagger-\mu^\dagger)(y^\dagger-\mu^\dagger)^\top](\Phi-\mathscr{I}) + \mathbb{E}_\theta[(y^*-\mu^*)]b^\top\}\iota \\
&= X^\top T_1 \Phi^{-1}\mathscr{M}^{-1} D(\mathscr{I}+D)\{V^*(D-\mathscr{I}) + C(\Phi-\mathscr{I})\}\iota.
\end{aligned}$$

Here, $V^* = \operatorname{diag}(\upsilon_1^*, \ldots, \upsilon_n^*)$ and $C = \operatorname{diag}(c_1, \ldots, c_n)$. Since y_t and y_u, for $t \neq u$, are independent, and $\mathbb{E}_{\theta_1}(y_t^* - \mu_t^{*(1)}) = 0$, we have $\mathbb{E}_{\theta_1}[(y_t^* - \mu_t^{*(1)})(y_u^* - \mu_u^*)] = 0$. Also, $\mathbb{E}_{\theta_1}[(y_t^* - \mu_t^{*(1)})(y_t^* - \mu_t^*)] = \mathbb{E}_{\theta_1}[(y_t^* - \mu_t^{*(1)})(y_t^* - \mu_t^{*(1)})] + \mathbb{E}_{\theta_1}[(y_t^* - \mu_t^{*(1)})(\mu_t^{*(1)} - \mu_t^*)] = \mathbb{E}_{\theta_1}[(y_t^* - \mu_t^{*(1)})^2] = \upsilon_t^{*(1)}$. Here, the superscript '(1)' indicates evaluation at θ_1.

After some algebra, we obtain

$$\mathbb{E}_{\theta_1}[U_\beta(\theta_1)\ell(\theta)] = X^\top T_1^{(1)} \Phi^{-1(1)}\mathscr{M}^{-1(1)} D^{(1)}(\mathscr{I}+D^{(1)})\{V^{*(1)}(D^{(1)}-\mathscr{I}) + C^{(1)}(\Phi-\mathscr{I})\}\iota.$$

Hence,

$$\begin{aligned}
\mathbb{E}_{\theta_1}[U_\beta(\theta_1)\ell(\theta_1)] - \mathbb{E}_{\theta_1}[U_\beta(\theta_1)\ell(\theta)] &= X^\top T_1^{(1)} \Phi^{-1(1)}\mathscr{M}^{-1(1)} D^{(1)}(\mathscr{I}+D^{(1)})\{V^{*(1)}(D^{(1)}-D) \\
&\quad + C^{(1)}(\Phi^{(1)}-\Phi)\}\iota.
\end{aligned}$$

Similarly, using (2.7) and (2.11) we obtain

$$\begin{aligned}
\mathbb{E}_\theta[U_\delta(\theta)\ell(\theta)] &= \mathbb{E}_\theta[H^\top T_2[P(y^*-\mu^*) + (y^\dagger - \mu^\dagger)]\{(y^*-\mu^*)^\top(D-\mathscr{I}) \\
&\quad + (y^\dagger-\mu^\dagger)^\top(\Phi-\mathscr{I}) + b^\top\}\iota] \\
&= H^\top T_2\{P\mathbb{E}_\theta[(y^*-\mu^*)(y^*-\mu^*)^\top](D-\mathscr{I}) + P\mathbb{E}_\theta[(y^*-\mu^*)(y^\dagger-\mu^\dagger)^\top] \\
&\quad \times (\Phi-\mathscr{I}) + \mathbb{E}_\theta[(y^\dagger-\mu^\dagger)(y^*-\mu^*)^\top](D-\mathscr{I}) + \mathbb{E}_\theta[(y^\dagger-\mu^\dagger)(y^\dagger-\mu^\dagger)^\top] \\
&\quad \times (\Phi-\mathscr{I})\}\iota \\
&= H^\top T_2\{PV^*(D-\mathscr{I}) + PC(\Phi-\mathscr{I}) + C(D-\mathscr{I}) + V^\dagger(\Phi-\mathscr{I})\}\iota.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}_{\theta_1}[U_\delta(\theta_1)\ell(\theta_1)] - \mathbb{E}_{\theta_1}[U_\delta(\theta_1)\ell(\theta)] &= H^\top T_2^{(1)}\{(P^{(1)}V^{*(1)} + C^{(1)})(D^{(1)}-D) \\
&\quad + (P^{(1)}C^{(1)} + V^{\dagger(1)})(\Phi^{(1)}-\Phi)\}\iota,
\end{aligned}$$

where $V^\dagger = \mathrm{diag}(\upsilon_1^\dagger, \ldots, \upsilon_n^\dagger)$.

We shall now move to the derivation of $\bar{\Upsilon}$ which is obtained from

$$\Upsilon = \begin{bmatrix} \mathrm{I\!E}_{\theta_1}[U_\beta(\theta_1)U_\beta^\top(\theta)] & \mathrm{I\!E}_{\theta_1}[U_\beta(\theta_1)U_\delta^\top(\theta)] \\ \mathrm{I\!E}_{\theta_1}[U_\delta(\theta_1)U_\beta^\top(\theta)] & \mathrm{I\!E}_{\theta_1}[U_\delta(\theta_1)U_\delta^\top(\theta)] \end{bmatrix}.$$

It follows from Equations (2.10) and (2.11) that

$$\begin{aligned}
\mathrm{I\!E}_{\theta_1}[U_\beta(\theta_1)U_\beta^\top(\theta)] &= \mathrm{I\!E}_{\theta_1}[X^\top T_1^{(1)}\Phi^{-1(1)}\mathscr{M}^{-1(1)}D^{(1)}(\mathscr{I}+D^{(1)})(y^*-\mu^{*(1)}) \\
&\quad \times (y^*-\mu^*)^\top(\mathscr{I}+D)D\mathscr{M}^{-1}\Phi^{-1}T_1X] \\
&= X^\top T_1^{(1)}\Phi^{-1(1)}\mathscr{M}^{-1(1)}D^{(1)}(\mathscr{I}+D^{(1)})V^{*(1)}(\mathscr{I}+D)D\mathscr{M}^{-1}\Phi^{-1}T_1X, \\
\mathrm{I\!E}_{\theta_1}[U_\delta(\theta_1)U_\delta^\top(\theta)] &= \mathrm{I\!E}_{\theta_1}[H^\top T_2^{(1)}\{P^{(1)}(y^*-\mu^{*(1)})+(y^\dagger-\mu^{\dagger(1)})\}\{(y^*-\mu^*)^\top P^\top \\
&\quad + (y^\dagger-\mu^\dagger)^\top\}T_2H] \\
&= H^\top T_2^{(1)}\{P^{(1)}V^{*(1)}P^\top + (P^{(1)}+P^\top)C + V^\dagger\}T_2H, \\
\mathrm{I\!E}_{\theta_1}[U_\beta(\theta_1)U_\delta^\top(\theta)] &= \mathrm{I\!E}_{\theta_1}[X^\top T_1^{(1)}\Phi^{-1(1)}\mathscr{M}^{-1(1)}D^{(1)}(\mathscr{I}+D^{(1)})(y^*-\mu^{*(1)}) \\
&\quad \times \{(y^*-\mu^*)^\top P^\top + (y^\dagger-\mu^\dagger)^\top\}T_2H] \\
&= X^\top T_1^{(1)}\Phi^{-1(1)}\mathscr{M}^{-1(1)}D^{(1)}(\mathscr{I}+D^{(1)})\{V^*P^\top + C^{(1)}\}T_2H, \\
\mathrm{I\!E}_{\theta_1}[U_\delta(\theta_1)U_\beta^\top(\theta)] &= \mathrm{I\!E}_{\theta_1}[H^\top T_2^{(1)}\{P^{(1)}(y^*-\mu^{*(1)})+(y^\dagger-\mu^{\dagger(1)})(y^*-\mu^*)^\top(\mathscr{I}+D) \\
&\quad \times D\mathscr{M}^{-1}\Phi^{-1}T_1X] \\
&= H^\top T_2^{(1)}\{P^{(1)}V^{*(1)}+C^{(1)}\}(\mathscr{I}+D)D\mathscr{M}^{-1}\Phi^{-1}T_1X.
\end{aligned}$$

Finally, by combining the results presented above it is possible to write $\bar{q}$ and $\bar{\Upsilon}$ as

$$\bar{q} = \begin{bmatrix} X^\top\hat{T}_1\hat{\Phi}^{-1}\hat{\mathscr{M}}^{-1}\hat{D}(\mathscr{I}+\hat{D})\{\hat{V}^*(\hat{D}-\tilde{D})+\hat{C}(\hat{\Phi}-\tilde{\Phi})\}\iota \\ H^\top\hat{T}_2\{(\hat{P}\hat{V}^*+\hat{C})(\hat{D}-\tilde{D})+(\hat{P}\hat{C}+\hat{V}^\dagger)(\hat{\Phi}-\tilde{\Phi})\}\iota \end{bmatrix}$$

and

$$\bar{\Upsilon} = \begin{bmatrix} \begin{array}{c} X^\top\hat{T}_1\hat{\Phi}^{-1}\hat{\mathscr{M}}^{-1}\hat{D}(\mathscr{I}+\hat{D})\hat{V}^* \\ \times(\mathscr{I}+\tilde{D})\tilde{D}\tilde{\mathscr{M}}^{-1}\tilde{\Phi}^{-1}\tilde{T}_1X \end{array} & \begin{array}{c} X^\top\hat{T}_1\hat{\Phi}^{-1}\hat{\mathscr{M}}^{-1}\hat{D}(\mathscr{I}+\hat{D}) \\ \times\{\hat{V}^*\tilde{P}+\hat{C}\}\tilde{T}_2H \end{array} \\ H^\top\hat{T}_2\{\hat{P}\hat{V}^*+\hat{C}\}(\mathscr{I}+\tilde{D})\tilde{D}\tilde{\mathscr{M}}^{-1}\tilde{\Phi}^{-1}\tilde{T}_1X & H^\top\hat{T}_2\{\hat{P}\hat{V}^*\tilde{P}+(\hat{P}+\tilde{P})\tilde{C}+\hat{V}^\dagger\}\tilde{T}_2H \end{bmatrix},$$

We emphasize that the two adjusted likelihood ratio statistics can be easily computed using standard software and computing environments since such computation only entail basic matrix operations.

CHAPTER 4

Simulation results

In this chapter we shall present the results of a set of Monte Carlo simulations that were performed to evaluate the finite sample performances of the standard likelihood ratio test (w) and its two corrected counterparts (w^* and w^{**}) in unit gamma regression models. Parameter estimation was carried out by numerically maximizing the log-likelihood function using the BFGS quasi-Newton optimization algorithm with analytic first derivatives. The number of Monte Carlo replications is $10,000$, i.e., all reported results are based on $10,000$ samples. All simulations were performed using the OX matrix programming language (Doornik, 2009).

Data generation is performed as follows. We generate a sample of size n from the gamma distribution with parameters μ and ϕ, and we then exponentiate the negative of such values. To see why such a data generating mechanism is valid, let $z \sim Gamma(\mu,\phi)$ and $y = \exp(-z)$, and denote by F_z and F_y (f_z and f_y) the corresponding distribution (density) functions. Notice that

$$F_y(a) = \Pr(y \leq a) = \Pr(\exp(-z) \leq a) = \Pr(z > -\log(a)) = 1 - F_z(-\log(a))$$

which implies

$$f_y(a) = -f_z(-\log(a))\left(-\frac{1}{a}\right) = \frac{\mu^{\phi}}{\Gamma(\phi)} a^{(\mu-1)}(-\log(a))^{(\phi-1)}, \quad 0 < y < 1.$$

Hence, the data generating scheme we use is valid. Under fixed dispersion, we use the above mechanism with μ replaced by μ_i, and under variable dispersion we, additionally, replace ϕ with ϕ_i, $i = 1,\ldots,n$.

4.1 Fixed dispersion unit gamma regression

At the outset we consider a fixed dispersion unit gamma regression model with mean submodel given by

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_{i4} x_{i4}, \tag{4.1}$$

$i = 1,\ldots,n$. The covariates values are obtained as random draws from the standard uniform distribution, i.e., from $\mathscr{U}(0,1)$. We shall consider two separate scenarios. In the first scenario, we test $\mathscr{H}_0 : \beta_4 = 0$. Hence, $l = 1$ (one restriction). The parameter values are $\beta_1 = -1.5$, $\beta_2 = -1.5$, $\beta_3 = 1.2$ and $\beta_4 = 0$. In the second scenario, we test $\mathscr{H}_0 : \beta_3 = \beta_4 = 0$. Hence, $l = 2$ (two restrictions). The parameter values were specified as $\beta_1 = 1$, $\beta_2 = 5$ and $\beta_3 = \beta_4 = 0$.

In both scenarios the null hypotheses were tested against bilateral alternative hypotheses. We shall consider three separate range of values for μ: close to zero (μ_1), around 0.5 (μ_2) and close to one (μ_3). We consider three values for the precision parameter, namely: $\phi = 5, 10, 30$. The sample sizes are $n = 20, 40, 60$. The tests significance levels are $\alpha = 10\%$, $\alpha = 5\%$ and $\alpha = 1\%$.

The tests null rejection rates are presented in Table 4.1 (entries are percentages). The reported results lead to interesting conclusions. First, the likelihood ratio test is considerably oversized (liberal) when the sample size is small. For instance, when $n = 20$, $\phi = 10$ and $l = 2$, the test null rejection is in excess of 15% at the 10% significance level for all three ranges of mean values. Second, the two corrected tests (w^* and w^{**}) perform considerably better than the standard likelihood ratio test. Under the same conditions ($n = 20$, $\phi = 10$, $l = 2$, 10% significance level) their null rejection rates range from 9.7% to 10% (w^*) and from 10.1% to 10.3% (w^{**}) across the different ranges of mean values. Third, the two corrected tests behave similarly, w^* tending to be slightly more conservative than w^{**}. Overall, they perform very well, i.e., they are quite effective when it comes to controlling the frequency of type I error. Fourth, the range of variation of the mean responses do not noticeably affect the tests performances. Fifth, the impact of the precision value on the tests finite sample performances is small.

The second set of simulations is based on the following fixed dispersion unit gamma regression model:

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1 + \sum_{j=2}^{p} \beta_j x_{ij}.$$

The interest lies in testing $\mathscr{H}_0 : \beta_2 = 0$ against $\mathscr{H}_1 : \beta_2 \neq 0$. We shall report numerical results for $p = 3, 4, 5$. Notice that the number of nuisance parameters increase with p. The sample sizes are $n = 20, 40, 60$, $\phi = 30$ and, as before, we consider the ranges of values for μ. The tests null rejection rates can be found in Table 4.2 (entries are percentages). The figures in this table show that the likelihood ratio test tends to become more liberal as the number of nuisance parameters increases. Consider, e.g., $n = 20$ and μ_2 (i.e., response means assuming values in the middle of the standard unit interval). At the 5% significance level, the test null rejection rates corresponding to $p = 3, 4, 5$ are, respectively, 7.0%, 8.4% and 9.4%. In contrast, the finite sample behaviors of the two corrected tests are much less affected by the increase in the number of nuisance parameters. Under the same conditions, for instance, the null rejection rates of w^* (w^{**}) are 4.7%, 4.9% and 5.1%(4.9%, 5.1% and 5.5%). It is also noteworthy that once again the two corrected tests behave similarly, w^* being slightly more conservative than w^{**}. As in the previous set of simulations, the range of variation of the mean responses has no noticeable impact on the tests performances.

Figure 4.1 contains quantile-quantile (QQ) plots for the three test statistics. The sample size is $n = 20$, $\phi = 5, 10$ and the response means assume values in the middle of the standard unit interval. In each panel we plot the test statistic empirical quantiles against the corresponding asymptotic quantiles, i.e., against χ^2_1 quantiles. There is good agreement between exact and asymptotic null distributions when the curve is close to the 45 degree line (the thin solid, diagonal line). Visual inspection of the QQ plots shows that the null distributions of the two corrected test statistics are much better approximated by the reference χ^2 square distribution than that of the likelihood ratio test statistic. It is also clear that the corrected test statistics are

Table 4.1 Null rejection rates (%), fixed dispersion I.

ϕ	l	n	μ_1 w	μ_1 w^*	μ_1 w^{**}	μ_2 w	μ_2 w^*	μ_2 w^{**}	μ_3 w	μ_3 w^*	μ_3 w^{**}
$\alpha = 10\%$											
30	1	20	14.9	10.0	10.3	14.7	10.0	10.4	14.7	10.0	10.4
		40	12.2	9.8	9.9	11.8	9.8	9.9	12.1	9.9	10.0
		60	11.3	9.9	9.9	11.3	9.8	9.9	11.1	9.8	9.9
	2	20	16.2	10.3	10.7	16.3	10.5	10.8	16.2	10.4	10.8
		40	13.1	10.0	10.2	13.0	10.2	10.3	12.9	10.2	10.3
		60	11.8	10.1	10.1	11.8	10.1	10.2	11.8	10.1	10.2
10	1	20	14.9	9.9	10.3	14.8	9.9	10.3	14.7	9.7	10.1
		40	12.0	9.9	9.9	11.7	9.6	9.7	11.9	9.7	11.9
		60	11.9	10.4	10.5	11.8	10.6	10.7	11.7	10.5	10.5
	2	20	15.6	9.7	10.1	15.8	10.0	10.3	15.5	9.9	10.1
		40	12.6	9.8	9.9	12.3	9.7	9.7	12.3	9.6	9.7
		60	11.8	10.1	10.1	11.5	10.0	10.0	11.4	10.1	10.1
5	1	20	15.0	10.4	10.7	15.0	10.4	10.8	14.7	10.3	10.6
		40	11.6	9.3	9.5	11.7	9.7	9.8	11.2	9.3	9.4
		60	11.3	9.8	10.0	11.4	10.0	10.1	11.2	9.8	9.9
	2	20	16.5	9.8	10.2	15.8	9.6	10.0	15.7	9.8	10.1
		40	12.7	10.0	10.0	12.6	10.0	10.0	12.3	9.9	10.0
		60	11.2	9.5	9.6	11.5	9.9	9.9	11.2	9.9	9.9
$\alpha = 5\%$											
30	1	20	8.4	5.2	5.4	8.6	5.2	5.4	8.2	5.1	5.3
		40	6.2	4.7	4.9	6.4	4.8	4.9	6.3	4.7	4.8
		60	6.0	5.2	5.2	6.2	5.1	5.2	5.8	5.1	5.2
	2	20	9.5	5.0	5.2	9.5	5.3	5.4	9.5	5.1	5.4
		40	6.7	5.1	5.2	6.7	4.7	4.8	6.7	4.8	4.9
		60	6.4	5.0	5.0	6.2	5.0	5.0	6.1	5.1	5.1
10	1	20	8.3	5.3	5.5	8.4	5.1	5.3	8.1	5.1	5.4
		40	6.4	4.8	4.9	6.0	4.6	4.7	6.1	4.6	4.7
		60	6.3	5.3	5.4	6.0	5.1	5.2	6.2	5.3	5.3
	2	20	9.2	5.2	5.4	9.1	5.1	5.3	8.9	5.2	5.3
		40	6.6	4.8	4.8	6.5	4.9	4.9	6.4	4.7	4.8
		60	6.4	5.0	5.1	6.2	5.0	5.0	6.3	5.2	5.2
5	1	20	8.9	5.3	5.6	8.5	5.1	5.4	8.5	5.1	5.3
		40	6.3	5.0	5.1	6.2	4.8	4.9	6.1	4.9	4.9
		60	5.6	4.8	4.9	5.9	5.0	5.2	5.6	4.8	4.8
	2	20	9.4	4.6	4.7	9.0	4.8	5.1	8.8	4.8	4.9
		40	6.9	4.8	4.8	6.6	4.9	4.9	6.6	4.8	4.8
		60	6.0	4.7	4.8	5.9	4.7	4.7	5.9	4.8	4.8
$\alpha = 1\%$											
30	1	20	2.4	1.0	1.1	2.4	1.0	1.0	2.3	0.9	1.0
		40	1.3	0.9	0.9	1.4	1.0	1.0	1.4	0.9	1.0
		60	1.4	1.0	1.0	1.4	1.0	1.0	1.3	1.0	1.1
	2	20	2.5	1.0	1.1	2.6	0.9	1.0	2.5	0.9	0.9
		40	1.6	0.9	1.0	1.5	1.0	1.0	1.5	1.0	1.0
		60	1.3	1.0	1.0	1.4	1.0	1.0	1.4	1.0	1.0
10	1	20	2.4	1.0	1.1	2.4	1.0	1.2	2.3	1.0	1.2
		40	1.4	0.9	1.0	1.3	0.9	0.9	1.4	0.9	0.9
		60	1.4	1.0	1.1	1.2	0.9	1.0	1.3	0.9	1.0
	2	20	2.8	1.0	1.1	2.7	1.0	1.1	2.7	1.0	1.1
		40	1.5	0.9	0.9	1.5	1.0	1.0	1.4	0.9	0.9
		60	1.4	1.0	1.0	1.5	1.0	1.0	1.4	1.1	1.1
5	1	20	2.4	1.0	1.1	2.4	1.0	1.1	2.2	1.0	1.1
		40	1.4	0.9	1.0	1.4	0.9	1.0	1.4	0.9	1.0
		60	1.3	1.0	1.1	1.3	1.0	1.1	1.4	1.0	1.0
	2	20	2.6	0.9	1.0	2.5	1.0	1.0	2.4	0.9	1.0
		40	1.4	0.8	0.8	1.5	0.8	0.9	1.3	0.8	0.9
		60	1.4	0.9	0.9	1.2	0.9	0.9	1.3	0.9	0.9

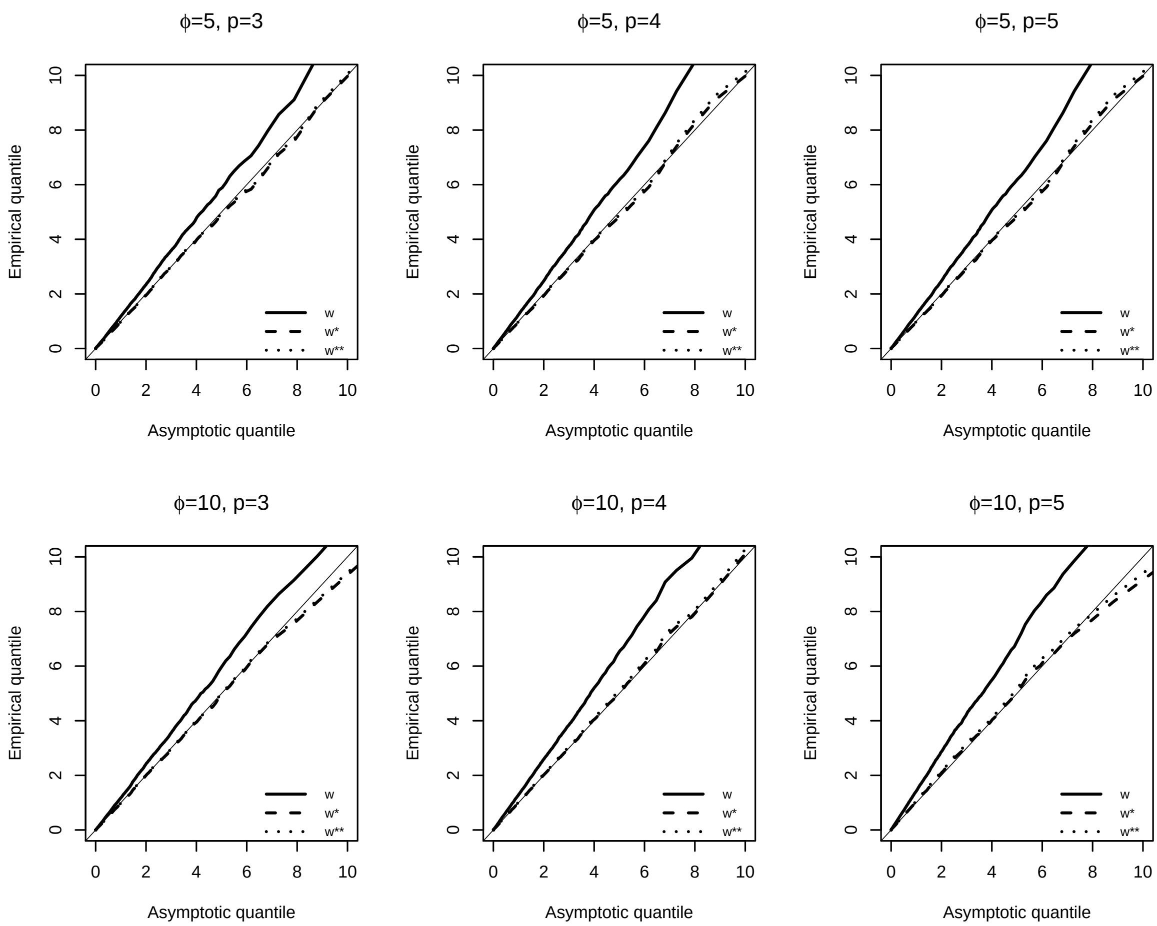

Figure 4.1 Quantile-quantile (QQ) plots, fixed dispersion, $l = 1$, $n = 20$.

Table 4.2 Null rejection rates (%), fixed dispersion II.

	$\alpha = 10\%$																										
	w									w^*									w^{**}								
$\phi = 30$	μ_1			μ_2			μ_3			μ_1			μ_2			μ_3			μ_1			μ_2			μ_3		
p	3	4	5	3	4	5	3	4	5	3	4	5	3	4	5	3	4	5	3	4	5	3	4	5	3	4	5
n																											
20	13.0	15.0	15.7	12.9	14.9	15.8	13.1	14.8	15.7	9.5	10.0	10.1	9.5	9.8	9.9	9.4	10.0	9.9	9.6	10.4	10.6	9.7	10.1	10.3	9.6	10.3	10.4
40	11.3	12.8	13.0	11.1	12.6	12.8	11.1	12.6	13.1	9.5	10.6	10.2	9.7	10.4	10.3	9.5	10.3	10.3	9.6	10.7	10.4	9.8	10.5	10.4	9.7	10.3	10.5
60	11.1	11.3	12.0	10.6	11.3	12.1	10.8	11.3	12.0	9.9	9.8	10.2	9.7	10.0	10.5	9.7	10.0	10.2	9.9	9.9	10.4	9.8	10.1	10.5	9.8	10.0	10.3
	$\alpha = 5\%$																										
20	7.2	8.7	9.7	7.0	8.4	9.4	7.0	8.4	9.3	4.8	5.0	5.1	4.7	4.9	5.1	4.7	4.9	4.9	4.9	5.2	5.4	4.9	5.1	5.5	4.8	5.1	5.3
40	6.0	6.9	7.2	5.9	6.7	7.1	6.0	6.8	7.3	4.9	5.4	5.3	4.8	5.4	5.4	4.8	5.2	5.4	5.0	5.4	5.4	4.8	5.4	5.4	4.9	5.3	5.6
60	5.4	5.8	6.4	5.6	5.8	6.4	5.5	5.7	6.3	4.8	4.8	5.1	4.9	4.9	5.0	4.7	4.9	5.0	4.9	4.9	5.2	4.9	4.9	5.1	4.8	5.0	5.0
	$\alpha = 1\%$																										
20	1.8	2.2	2.9	1.8	2.1	2.7	1.8	2.2	2.7	0.9	1.0	0.9	0.8	0.9	0.9	0.8	0.8	1.0	1.0	1.1	1.0	0.9	1.0	1.1	0.9	0.9	1.1
40	1.4	1.7	1.9	1.4	1.7	1.8	1.4	1.6	1.8	1.0	1.1	1.1	1.0	1.1	1.1	1.0	1.0	1.1	1.1	1.1	1.1	1.0	1.1	1.1	1.1	1.1	1.2
60	1.2	1.3	1.5	1.2	1.3	1.4	1.2	1.3	1.3	0.9	1.0	1.0	1.0	1.0	0.9	0.9	0.9	1.0	1.0	1.1	1.1	1.0	1.0	1.0	1.0	0.9	1.0

less affected by the increase in the number of nuisance parameters.

4.2 Variable dispersion unit gamma regression

Next, we consider the variable precision unit gamma regression model given by

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_{i4} x_{i4}$$

and

$$\log(\phi_i) = \delta_1 + \delta_2 h_{i2} + \delta_3 h_{i3} + \delta_4 h_{i4},$$

$i = 1,\dots,n$. Again, the covariates values are obtained as random standard uniform draws. We shall consider two separate scenarios. In the first scenario, we test $\mathscr{H}_0 : \delta_2 = \delta_3 = \delta_4 = 0$ $(l = 3)$ against a two-sided alternative hypothesis. Notice that we test the null hypothesis of constant dispersion against the alternative hypothesis of variable dispersion. The parameter values used for data generation are $\beta_1 = 1.5$, $\beta_2 = 1.5$, $\beta_3 = 4.5$, $\beta_4 = -3.5$, $\delta_1 = \log(30)$, $\delta_2 = \delta_3 = \delta_4 = 0$, $\delta_1 = \log(30), \log(10), \log(5)$. In the second scenario, we test $\mathscr{H}_0 : \beta_3 = \beta_4 = 0, \delta_2 = \delta_3 = \delta_4 = 0$ $(l = 5)$. Data generation was carried out using $\beta_1 = 1.5$, $\beta_2 = 1.5$, $\beta_3 = 0$, $\beta_4 = 0$, $\delta_1 = (\log(30), \log(10), \log(5))$, $\delta_2 = \delta_3 = \delta_4 = 0$. The sample sizes are $n = 20, 30, 40, 50, 60$ in the first scenario and $n = 20, 40, 60$ in the second scenario. In both scenarios, the tests significance levels are $\alpha = 10\%$, $\alpha = 5\%$ and $\alpha = 1\%$. The results (tests null rejection rates) are presented in Table 4.3 (entries are percentages).

Table 4.3 Null rejection rates (%), variable dispersion I.

		$\alpha = 10\%$			$\alpha = 5\%$			$\alpha = 1\%$		
δ_1	n	w	w^*	w^{**}	w	w^*	w^{**}	w	w^*	w^{**}
log(30)	20	34.8	12.5	15.6	24.5	7.0	9.0	10.8	1.6	2.3
	30	22.5	10.6	11.8	14.2	5.8	6.4	4.8	1.1	1.2
	40	17.6	9.7	10.2	10.4	5.0	5.3	2.9	1.0	1.0
	50	15.7	9.9	10.2	8.6	4.7	4.9	2.3	0.9	1.0
	60	14.6	10.2	10.3	8.0	5.1	5.2	2.1	1.0	1.1
log(10)	20	33.2	12.9	15.7	23.5	7.2	9.1	10.2	1.7	2.4
	30	22.4	11.4	12.3	14.5	5.8	6.5	4.5	1.2	1.4
	40	17.6	10.7	11.1	10.8	5.2	5.4	3.2	1.2	1.2
	50	15.1	10.1	10.2	8.6	5.1	5.2	2.3	1.0	1.1
	60	13.9	9.6	9.7	7.7	4.8	4.8	1.8	0.8	0.9
log(5)	20	32.5	12.5	15.6	22.7	7.1	9.2	10.0	2.1	2.7
	30	22.2	11.6	12.6	14.0	6.1	6.6	4.7	1.4	1.6
	40	17.7	10.8	11.2	10.4	5.5	5.7	2.8	1.0	1.1
	50	14.7	9.7	9.9	8.4	4.9	5.0	2.0	1.0	1.1
	60	14.4	10.5	10.6	8.2	5.5	5.6	2.2	1.1	1.2

The Table 4.3 show that the likelihood ratio test tends to be considerably liberal in small samples, i.e., it tends to overreject the null hypothesis when such a hypothesis is true. For instance, when $n = 20$, $\delta_1 = \log(10)$ and $\alpha = 0.10$, the test null rejection rate equals 33.2%, i.e., it is over three times larger than the test significance level. When $n = 50$, it equals 15.1%; it still is over 50% larger than the nominal level. The corrected tests are less size distorted. Under the same conditions, e.g., the null rejection rates of w^* (w^{**}) are 12.9% and 15.7% for $n = 20$. It is noteworthy that w^* is less oversized than w^{**}, especially when n is small.

The tests null rejection rates for the second scenario (joint test on the parameters of both submodels) are presented in Table 4.4 (entries are percentages). Once again, the likelihood ratio test is substantially oversized in small samples. For instance, when $n = 20$, $\delta_1 = \log(30)$ and $\alpha = 10\%$, its null rejection rate equals 34.9%. The two corrected tests perform well. Under the same circumstance, their null rejection rates are 8.8% (w^*) and 10.5% (w^{**}). As expected, the tests performances improve as the sample size increases.

Table 4.4 Null rejection rates (%), variable dispersion II.

		$\alpha = 10\%$			$\alpha = 5\%$			$\alpha = 1\%$		
δ_1	n	w	w^*	w^{**}	w	w^*	w^{**}	w	w^*	w^{**}
	20	34.9	8.8	10.5	24.6	5.5	6.5	10.3	2.1	2.6
log(30)	40	17.7	8.9	9.4	10.3	4.1	4.4	3.0	0.8	0.9
	60	14.7	9.4	9.7	7.9	4.4	4.6	1.8	0.9	0.9
	20	34.1	8.2	9.9	23.8	4.8	5.8	9.8	1.8	2.3
log(10)	40	17.5	8.7	9.3	10.3	4.4	4.7	2.8	0.9	1.0
	60	13.9	9.1	9.2	7.8	4.6	4.7	1.9	0.9	0.9
	20	34.1	8.2	9.9	23.8	4.8	5.8	9.8	1.8	2.3
log(5)	40	17.5	8.7	9.3	10.3	4.4	4.7	2.8	0.9	1.0
	60	13.9	9.1	9.2	7.8	4.6	4.7	1.9	0.9	0.9

Figure 4.2 contains QQ plots for the second simulation design. It is clear that the null distribution of w is poorly approximated by the limiting chi-squared distribution. It is also clear that such an approximation is much more precise when used with the two corrected test statistics, especially when $n \geq 40$. Notice that the case where there are only twenty observations in the sample is quite challenging since there are five restrictions under test. Even with $n = 20$, however, the chi-squared approximation is somewhat precise when used with w^* and w^{**}, except in the distribution upper tail.

Overall, the numerical evidence in this chapter shows that likelihood ratio testing inferences in unit gamma regressions can be quite unreliable in small samples. Much more precises can be achieved by using the two corrected test statistics derived in this dissertation. We also note that testing inferences based on w^* tend to be slightly more accurate than those based on w^{**}.

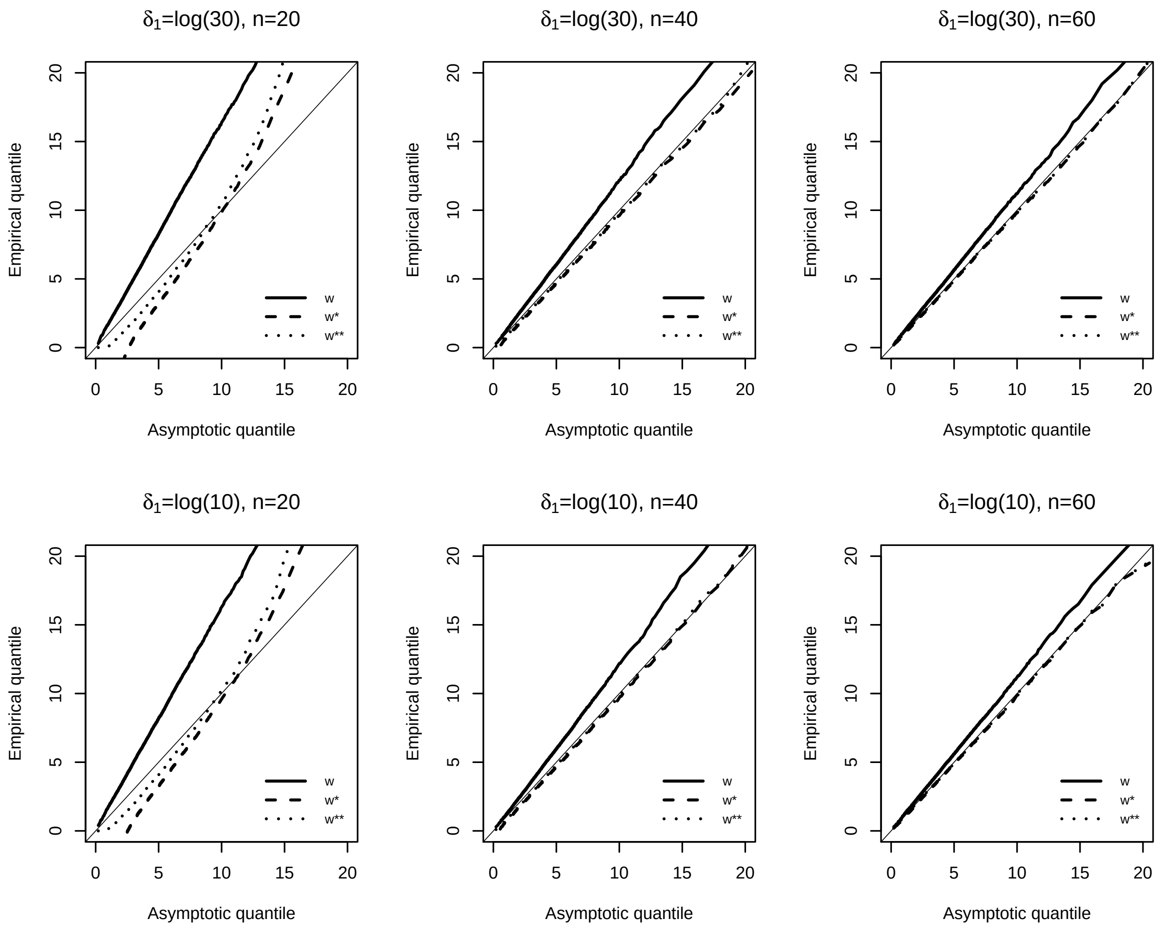

Figure 4.2 Quantile-quantile (QQ) plots, variable dispersion, $l = 5$.

CHAPTER 5

An empirical application

In what follows we shall present an empirical application that uses both the unit gamma and the beta regression model. We shall use a dataset analyzed by Smithson and Verkuilen (2006) that contains 44 observations on reading accuracy of dyslexic and nondyslexic Australian children; the data are presented in Table A.1. The variable of interest (y) are reading accuracy indices of such children. The independent variables are: dyslexia versus non-dyslexia status (x_2) and nonverbal IQ converted to z-scores (x_3). We also consider two additional covariates, namely: an interaction variable ($x_2 \times x_3$) and z-scores squared (x_3^2). The participants (19 dyslexics and 25 controls) were students from primary schools in the Australian Capital Territory. The ages of the 44 children range from eight years five months to twelve years three months. The covariate x_2 is a dummy variable, which equals 1 if the child is dyslexic and -1 otherwise. The observed scores were linearly transformed from their original scale to the open unit interval $(0,1)$; see Smithson and Verkuilen (2006). These dara were also analyzed by Espinheira et al. (2008), Grün et al. (2011), Pinto Ferreira de Queiroz (2011), Cribari-Neto and Queiroz (2014) and Bayer and Cribari-Neto (2015) using beta regression models. It is noteworthy that different regression models were used in the literature. One of such models uses the interation variable ($x_2 \times x_3$) and the squared values of the z-scores in the precision submodel. According to Bayer and Cribari-Neto (2015), this is the model selected by most standard beta regression model selection approaches. We shall denote such a model by BetaModel–I. It is given by

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3}^2 + \beta_4 (x_{i2} \times x_{i3}^2) \tag{5.1}$$

and

$$\log(\phi_i) = \delta_1 + \delta_2 x_{i2} + \delta_3 x_{i3} + \delta_4 x_{i3}^2 + \delta_5 (x_{i2} \times x_{i3}), \tag{5.2}$$

$i = 1, \ldots, 44$. Notice that here x_3^2 and $x_2 \times x_3$ are included as covariates in the precision submodel. We shall investigate whether such variables should be included in such a submodel. We shall consider both beta and unit gamma regression models. Since the sample size is small, all testing inferences shall be carried out at the 10% significance level.

5.1 Beta regression modeling

At the outset we assume that the response (y) is beta distributed. As explained above, the interest lies in determining whether x_3^2 and $x_2 \times x_3$ should be used as precision covariates in BetaModel–I. To that end, we test the null hypothesis $\mathscr{H}_0 : \delta_4 = \delta_5 = 0$ against a two-sided alternative hypothesis. The likelihood ratio test statistic (w) equals 16.410 (p-value < 0.001).

We also computed the adjusted test statistics w^* and w^{**} using the results in Ferrari and Pinheiro (2011); their values are, respectively, 5.544 (p-value $= 0.063$) and 7.342 (p-value $= 0.025$). All three tests reject the null hypothesis at the 10% significance level, our significance level of reference. We note, however, that only the standard likelihood ratio test rejects the null hypothesis at the 5% significance level; indeed, it does so even at the 1% significance level.

Next, we test $\mathscr{H}_0 : \delta_5 = 0$ against $\mathscr{H}_1 : \delta_5 \neq 0$. We obtain $w = 7.248$ (p-value $= 0.007$), $w^* = 1.067$ (p-value $= 0.302$) and $w^{**} = 2.385$ (p-value $= 0.123$). It is noteworthy that the testing inference based on the standard likelihood ratio test ($\mathscr{H}_0$ is rejected, even at the 5% significance level) is different from that obtained using the corrected tests ($\mathscr{H}_0$ is not rejected).

The maximum likelihood parameter estimates and the corresponding standard errors for BetaModel–I are given in Table 5.1.

Table 5.1 Parameter estimates and standard errors (S.E.): BetaModel–I.

Parameter	β_1	β_2	β_3	β_4	δ_1	δ_2	δ_3	δ_4	δ_5
Estimate	1.064	−0.856	0.448	−0.394	2.668	1.513	1.587	1.842	1.456
S.E.	0.150	0.149	0.056	0.056	0.306	0.270	0.257	0.290	0.360

Based on the testing inferences yielded by the two corrected tests, we removed the interaction variable ($x_2 \times x_3$) from the precision submodel, thus arriving at the following reduced model:

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3}^2 + \beta_4 (x_{i2} \times x_{i3}^2) \tag{5.3}$$

and

$$\log(\phi_i) = \delta_1 + \delta_2 x_{i2} + \delta_3 x_{i3} + \delta_4 x_{i3}^2, \tag{5.4}$$

$i = 1, \ldots, 44$. We shall denote such a model by BetaModel–II. Its parameter estimates and standard errors are presented in Table 5.2.

Table 5.2 Parameter estimates and standard errors: BetaModel–II.

Parameter	β_1	β_2	β_3	β_4	δ_1	δ_2	δ_3	δ_4
Estimate	1.116	−0.791	0.443	−0.411	2.593	1.249	1.050	0.897
S.E.	0.151	0.151	0.070	0.070	0.307	0.262	0.251	0.209

We now test $\mathscr{H}_0 : \delta_4 = 0$ against a two-sided alternative hypothesis. The test statistics values are $w = 9.162$ (p-value $= 0.002$), $w^* = 5.159$ (p-value $= 0.023$) and $w^{**} = 5.596$ (p-value $= 0.018$). The three tests reject the null hypothesis, thus suggesting that x_3^2 should be included as a precision covariate.

Finally, we test $\mathscr{H}_0 : \delta_2 = \delta_3 = \delta_4 = 0$. For this test, $w = 33.689$ (p-value < 0.001), $w^* = 20.549$ (p-value < 0.001) and $w^{**} = 21.830$ (p-value < 0.001). All three tests reject the null hypothesis, even at the 1% significance level. Hence, we consider BetaModel–II to be the correct model based on which all futher inferences should be made.

We computed the pseudo-R^2, AIC and BIC values for the two beta regression fitted models, i.e., for BetaModel–I and BetaModel–II; see Table 5.3. The pseudo-R^2 we compute is that of Nagelkerke (1991), i.e., $R^2 = 1-(L_{null}/L_{fit})^{2/n}$, where L_{null} is the maximized likelihood function using only the intercept (no regressors used) and L_{fit} is the maximized likelihood function based on the model at hand (regressors used); see Long (1997). Notice that both AIC and BIC favor BetaModel–I. However, since the interaction precision covariate is not statistically significance, we still select BetaModel–II as the best fitting model.

Table 5.3 Pseudo-R^2, AIC and BIC values; BetaModel–I and BetaModel–II .

Model	pseudo-R^2	AIC	BIC
BetaModel–I	0.628	−126.76	−110.70
BetaModel–II	0.624	−121.51	−107.24

5.2 Unit gamma regression modeling

We shall now turn to unit gamma regression modeling, i.e., we shall assume that the response (y) follows the unit gamma law. At the outset, we consider the model defined by (5.1) and (5.2), which we denote by GammaModel–I. We test the null hypothesis $\mathscr{H}_0 : \delta_4 = \delta_5 = 0$ against a two-sided alternative hypothesis. The values of the likelihood ratio test statistic (w) and its two corrected variants (w^* and w^{**}) are, respectively, 6.876 (p-value = 0.032), 0.380 (p-value = 0.827) and 1.914 (p-value = 0.384). It is noteworthy that the testing inferences are quite different when based on the standard likelihood ratio test ($\mathscr{H}_0$ is rejected) and on the corrected tests ($\mathscr{H}_0$ is not rejected). Indeed, the p-value of the w^* test, the best performing test in our Monte Carlo simulations, is quite large (in excess of 0.8), which indicates that there is very little evidence against the null hypothesis. The parameter estimates and standard errors for GammaModel–I can be found in Table 5.4.

Table 5.4 Parameter estimates and standard errors: GammaModel–I.

Parameter	β_1	β_2	β_3	β_4	δ_1	δ_2	δ_3	δ_4	δ_5
Estimate	1.099	−0.820	0.427	−0.413	1.290	2.209	1.103	1.122	1.027
S.E.	0.152	0.152	0.058	0.058	0.281	0.250	0.254	0.286	0.361

Based on the testing inference reached using the two corrected tests, we removed the covariates x_3^2 and $x_2 \times x_3$ from the precision submodel, thus arriving at the model we denote by GammaModel–II:

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3}^2 + \beta_4 (x_{i2} \times x_{i3}^2) \tag{5.5}$$

and

$$\log(\phi_i) = \delta_1 + \delta_2 x_{i2} + \delta_3 x_{i3}, \tag{5.6}$$

$i = 1, \ldots, 44$. The parameter estimates and standard errors obtained for such a model are displayed in Table 5.5.

Table 5.5 Parameter estimates and standard errors: GammaModel–II.

Parameter	β_1	β_2	β_3	β_4	δ_1	δ_2	δ_3
Estimate	1.062	−0.663	0.474	−0.476	1.603	2.030	0.625
S.E.	0.150	0.150	0.086	0.086	0.204	0.241	0.241

We now test the null hypothesis of fixed precision, i.e., we test $\mathscr{H}_0 : \delta_2 = \delta_3 = 0$. The values of the test statistics are $w = 47.920$, $w^* = 41.923$ and $w^{**} = 42.111$; all three p-values are quite small and, as a consequence, the null hypothesis is rejected by all three tests. There is, thus, evidence, of variable dispersion.

Table 5.6 contains the pseudo-R^2, AIC and BIC values for the two unit gamma regression models. The pseudo-R^2 and the AIC favor GammaModel–I whereas the BIC favors GammaModel–II. Since, the precision covariates x_3^2 and $x_2 \times x_3$ do not seem to be statistically significant according to the adjusted tests, we select GammaModel–II.

Table 5.6 Pseudo-R^2, AIC and BIC values; GammaModel–I and GammaModel–II.

Model	pseudo-R^2	AIC	BIC
GammaModel–I	0.627	−126.89	−110.83
GammaModel–II	0.607	−124.01	−111.52

It is noteworthy that the AIC and BIC values of the selected unit gamma regression model (−124.01 and −111.52) are smaller than those of the chosen beta regression model (−121.51 and −107.24); they are also slightly smaller than those of the competing beta regression model. It then follows that there is some evidence that the unit gamma regression model outperforms the beta regression model for the data at hand.

The above empirical application shows that it is important to perform accurate testing inferences when modeling data that assume values in the standard unit interval. When the sample size is small we recommend that such testing inferences be based on the two corrected tests developed in this dissertation, especially on w^*.

Table 5.7 P-values of the tests J.

	Statistic		
J test	w	w^*	w^{**}
GammaModel–II vs. BetaModel–II	0.0000	0.0866	0.0003
BetaModel–II vs. GammaModel–II	0.0000	0.0004	0.0004

Next, we wish to distinguish between the following models by means of a hypothesis test: GammaModel–II and BetaModel–II. Since such models are non-nested, we shall use the J test;

for details, see Cribari-Neto and Lucena (2015) and Cribari-Neto and Lucena (2017). *J* testing inferences were carried out using the standard likelihood ratio test statistic (w) and also de two corrected test statistics (w^* and w^{**}). The tests p-values are presented in Table 5.7. Based on the w^* test, we do not reject GammaModel–II and reject BetaModel–II at the 5% significance level. There is thus evidence that the unit gamma regression model yields a better fit than the beta regression model for these data.

CHAPTER 6
Concluding remarks

The unit gamma regression can be used to model responses that assume values in the standard unit interval, i.e., with dependent variables that assume values in $(0,1)$, such as rates and proportions. It is an alternative to the beta regression model. The unit gamma regression model is based on the assumption that the variable of interest follows the unit gamma law which is parameterized in terms of mean and precision parameters. Inference on the parameters that index the model is typically carried out using the likelihood ratio test. Such a test, however, tends to be quite inaccurate in small samples. In particular, it tends to be liberal, i.e., it overrejects the null hypothesis when such a hypothesis is true. In this dissertation, we derived two modified likelihood ratio tests statistics that are expected to deliver more accurate inferences when the sample size is small. They are obtained using a correction proposed by Skovgaard (2001). We considered both fixed and variable precision unit gamma regression models. The latter contains a submodel for the response mean and a separate submodel for the precision whereas in the former only the mean varies across observations.

We list the main contributions:

1. We derived two modified likelihood ratio test statistics, w^* and w^{**}, following the proposal made by Skovgaard (2001) for the class of unit gamma regression models. We obtained results for both fixed and variable precision models.

2. We performed Monte Carlo simulations to evaluate the tests finite sample behavior. The numerical evidence showed that the standard likelihood ratio testing inference can be quite inaccurate when the sample size is small. In particular, the test can be considerably liberal. The two corrected tests, in contrast, tend to be much less size distorted, thus yielding more accurate inferences. Our numerical results have also shown that one of the modified tests tends to be less liberal, i.e., more accurate: w^*.

3. We presented an empirical application in which the variable of interest was modeled using both a beta regression model and a unit gamma regression model. It is noteworthy that the corrected tests yielded an inference that is different from that obtained using the standard likelihood ratio test.

In short, we obtained two variants of the likelihood ratio test that are expected to deliver more accurate inferences in small samples. The numerical evidence we presented showed that the test we denoted by w^* is particularly accurate in small samples.

In future research we shall derive the Skovgaard adjustment (Skovgaard, 2001) for the classes of Kumaraswamy (Kumaraswamy, 1980) and simplex (Barndorff-Nielsen and Jørgensen, 1991; Jorgensen, 1997) regression models, which can also be used to model random variables that assume values in the standard unit interval.

References

M. Abdel-Fattah. Heteroscedastic generalized gamma regression for bounded continuous data. *Communications in Statistics – Theory and Methods*, forthcoming.

H. Akaike. Information theory and an extension of the maximum lilelihood principle. In *2nd International Symposium on Information Theory, Akademiai Kiado, Budapest, 1973*, pages 267–281, 1973.

H. Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.

H. Akaike. A bayesian analysis of the minimum aic procedure. *Annals of the Institute of Statistical mathematics*, 30(1):9–14, 1978.

O. Barndorff-Nielsen. On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, 70(2):343–365, 1983.

O. E. Barndorff-Nielsen. Infereni on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, 73(2):307–322, 1986.

O. E. Barndorff-Nielsen. Modified signed log likelihood ratio. *Biometrika*, 78(3):557–563, 1991.

O. E. Barndorff-Nielsen. Adjusted versions of profile likelihood and directed likelihood, and extended likelihood. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 125–140, 1994.

O. E. Barndorff-Nielsen and B. Jørgensen. Some parametric models on the simplex. *Journal of Multivariate Analysis*, 39(1):106–116, 1991.

M. S. Bartlett. Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, pages 268–282, 1937.

M. S. Bartlett and D. G. Kendall. The statistical analysis of variance-heterogeneity and the logarithmic transformation. *Supplement to the Journal of the Royal Statistical Society*, 8(1): 128–138, 1946.

F. M. Bayer and F. Cribari-Neto. Bootstrap-based model selection criteria for beta regressions. *Test*, 24(4):776–795, 2015.

J. Brehm and S. Gates. Donut shops and speed traps: Evaluating models of supervision on police behavior. *American Journal of Political Science*, pages 555–581, 1993.

G. M. Cordeiro and S. L. P. Ferrari. A modified score test statistic having chi-squared distribution to order n^{-1}. *Biometrika*, 78(3):573–582, 1991.

D. Cox and N. Reid. A note on the difference between profile and modified profile likelihood. *Biometrika*, 79(2):408–411, 1992.

D. R. Cox and N. Reid. Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–39, 1987.

F. Cribari-Neto and G. M. Cordeiro. On bartlett and bartlett-type corrections francisco cribari-neto. *Econometric reviews*, 15(4):339–367, 1996.

F. Cribari-Neto and S. E. F. Lucena. Non-nested hypothesis testing in the class of varying dispersion beta regressions. *Journal of Applied Statistics*, 42(5):967–985, 2015. doi: 10.1080/02664763.2014.993368. URL `http://dx.doi.org/10.1080/02664763.2014.993368`.

F. Cribari-Neto and S. E. F. Lucena. Non-nested hypothesis testing inference for gamlss models. *Journal of Statistical Computation and Simulation*, 87(6):1189–1205, 2017. doi: 10.1080/00949655.2016.1255946. URL `http://dx.doi.org/10.1080/00949655.2016.1255946`.

F. Cribari-Neto and M. P. Queiroz. On testing inference in beta regressions. *Journal of Statistical Computation and Simulation*, 84(1):186–203, 2014.

W. C. Davidon. Variable metric method for minimization.[in fortran for ibm 704]. Technical report, Argonne National Lab., IL (USA), 1966.

W. C. Davidon. Variable metric method for minimization. *SIAM Journal on Optimization*, 1 (1):1–17, 1991.

J. A. Doornik. *Ox: An Object-Oriented Matrix Programming Language*. Timberlake Consultants Press, London, 2009.

P. L. Espinheira, S. L. P. Ferrari, and F. Cribari-Neto. On beta regression residuals. *Journal of Applied Statistics*, 35(4):407–419, 2008.

S. Ferrari and F. Cribari-Neto. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7):799–815, 2004.

S. L. P. Ferrari and A. H. M. A. Cysneiros. Skovgaard's adjustment to likelihood ratio tests in exponential family nonlinear models. *Statistics & Probability Letters*, 78(17):3047–3055, 2008.

S. L. P. Ferrari and E. C. Pinheiro. Improved likelihood inference in beta regression. *Journal of Statistical Computation and Simulation*, 81(4):431–443, 2011.

A. C. Frery and F. Cribari-Neto. *Elementos de estatística computacional usando plataformas de software livre/gratuito*. IMPA, 2005.

A. Grassia. On a family of distributions with argument between 0 and 1 obtained by transformation of the gamma and derived compound distributions. *Australian & New Zealand Journal of Statistics*, 19(2):108–114, 1977.

B. Grün, I. Kosmidis, and A. Zeileis. Extended beta regression in r: shaken, stirred, mixed, and partitioned. Technical report, Working Papers in Economics and Statistics, 2011.

D. Hancox, C. J. Hoskin, and R. S. Wilson. Evening up the score: sexual selection favours both alternatives in the colour-polymorphic ornate rainbowfish. *Animal Behaviour*, 80(5): 845–851, 2010.

B. Jorgensen. *The theory of dispersion models*. CRC Press, 1997.

R. Kieschnick and B. D. McCullough. Regression analysis of variates observed on $(0,1)$: percentages, proportions and fractions. *Statistical Modelling*, 3(3):193–213, 2003.

P. Kumaraswamy. A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46(1-2):79–88, 1980.

D. N. Lawley. A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, 43(3/4):295–303, 1956.

E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer, 2nd edition, 2011.

J. S. Long. Regression models for categorical and limited dependent variables. *Advanced quantitative techniques in the social sciences*, 7, 1997.

P. McCullagh and J. A. Nelder. Generalized linear models, 1989.

P. McCullagh and R. Tibshirani. A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 325–344, 1990.

T. F. N. Melo, K. L. P. Vasconcellos, and A. J. Lemonte. Some restriction tests in a new class of regression models for proportions. *Computational Statistics & Data Analysis*, 53 (12):3972–3979, 2009.

A. M. Mousa, A. A. El-Sheikh, and M. A. Abdel-Fattah. A gamma regression for bounded continuous variables. *Advances and Applications in Statistics*, 49(4):305–326, 2016.

N. J. D. Nagelkerke. A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692, 1991.

J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, 2nd edition, 2006.

M. S. Oliveira. *Um Modelo de Regressão Beta: Teoria e Aplicações*. PhD thesis, Instituto de Matemática e Estatística da Universidade de São Paulo, 12/04/2004., 2004.

R. Ospina, F. Cribari-Neto, and K. L. P. Vasconcellos. Improved point and interval estimation for a beta regression model. *Computational Statistics & Data Analysis*, 51(2):960–981, 2006.

T. L. Pereira. *Regressão Beta Inflacionada: Inferência e Aplicações*. PhD thesis, PhD thesis, Universidade Federal de Pernambuco (UFPE), 2010.

T. L. Pereira and F. Cribari-Neto. Modified likelihood ratio statistics for inflated beta regressions. *Journal of Statistical Computation and Simulation*, 84(5):982–998, 2014.

M. Pinto Ferreira de Queiroz. Testes de hipóteses em regressão beta baseados em verossimilhança perfilada ajustada e em bootstrap. 2011.

G. Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

P. K. Sen, J. M. Singer, and A. C. P. De Lima. *From finite sample to asymptotic methods in statistics*, volume 28. Cambridge University Press, 2010.

A. B. Simas, W. Barreto-Souza, and A. V. Rocha. Improved estimators for a general class of beta regression models. *Computational Statistics and Data Analysis*, 54:348–366, 2010.

I. M. Skovgaard. An explicit large-deviation approximation to one-parameter tests. *Bernoulli*, pages 145–165, 1996.

I. M. Skovgaard. Likelihood asymptotics. *Scandinavian Journal of Statistics*, 28(1):3–32, 2001.

M. Smithson and J. Verkuilen. A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods*, 11(1):54, 2006.

S. E. Stern. A second-order adjustment to the profile likelihood in the case of a multidimensional parameter of interest. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(3):653–665, 1997.

R. Team. Rstudio: Integrated development for r. *RStudio, Inc., Boston, MA. URL: http://www.rstudio.com*, 2014.

G. Venter. Review of optimization techniques. *Encyclopedia of Aerospace Engineering*, 2010.

C. Zucco. The president's 'new' constituency: Lula and the pragmatic vote in brazil's 2006 presidential elections. *Journal of Latin American Studies*, pages 29–49, 2008.

APPENDIX A

Appendix A - Data

Table A.1: Reading accuracy data.

Observation	y	x_1	x_2
1	0.88386	−1	0.82700
2	0.76524	−1	0.59000
3	0.91508	−1	0.47100
4	0.98376	−1	1144
5	0.88386	−1	−0.67600
6	0.70905	−1	−0.79500
7	0.77148	−1	−0.28100
8	0.99000	−1	−0.91400
9	0.99000	−1	−0.04300
10	0.99000	−1	0.90700
11	0.99000	−1	0.51100
12	0.99000	−1	1223
13	0.99000	−1	0.5900
14	0.99000	−1	1856
15	0.99000	−1	−0.39900
16	0.99000	−1	0.59000
17	0.70281	−1	−0.04300
18	0.99000	−1	1738
19	0.66535	−1	0.47100
20	0.99000	−1	1619
21	0.95878	−1	1144
22	0.99000	−1	−0.20100
23	0.73402	−1	−0.28100
24	0.64662	−1	0.59000
25	0.99000	−1	1777
26	0.57794	1	−0.08300
27	0.64038	1	−0.16200
28	0.45932	1	−0.79500
29	0.65286	1	−0.28100
30	0.60916	1	−0.87400
31	0.60916	1	0.31300

(continues on the next page)

Table A.1 – continued from the previous page

Observation	y	x_1	x_2
32	0.54048	1	0.70900
33	0.57170	1	1223
34	0.70281	1	−1.23000
35	0.56546	1	−0.16200
36	0.53424	1	−0.99300
37	0.57794	1	−1191
38	0.69032	1	−1745
39	0.54673	1	−1745
40	0.68408	1	−0.43900
41	0.59043	1	−1666
42	0.62165	1	−1507
43	0.67159	1	−0.51800
44	0.66535	1	−1.27000