UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

CRISTINE RAUBER OLIVEIRA

CORRECTED LIKELIHOOD RATIO TEST STATISTICS FOR A CLASS OF BETA REGRESSIONS WITH PARAMETRIC LINK FUNCTION

Recife
2020

CRISTINE RAUBER OLIVEIRA

CORRECTED LIKELIHOOD RATIO TEST STATISTICS FOR A CLASS OF BETA REGRESSIONS WITH PARAMETRIC LINK FUNCTION

Dissertação apresentada ao Programa de Pós-Graduação em Estatística do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco, como requisito parcial à obtenção do título de Mestre em Estatística.
Área de Concentração: Estatística Matemática

Orientador: Prof. Francisco Cribari-Neto

Co-Orientador: Prof. Fábio Mariano Bayer

Recife

2020

Catalogação na fonte
Bibliotecária Arabelly Ascoli CRB4-2068

O48c Oliveira, Cristine Rauber
Corrected likelihood ratio test statistics for a class of beta regressions with parametric link function / Cristine Rauber Oliveira. – 2020.
62.: il., fig., tab.

Orientador: Francisco Cribari-Neto
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN. Estatística. Recife, 2020.
Inclui referências e apêndice.

1. Função de ligação. 2. Função de ligação paramétrica. 3. Regressão beta. 4. Teste da razão de verossimilhanças. I. Cribari-Neto, Francisco (orientador). II. Título.

519.5 CDD (22. ed.) UFPE-CCEN 2020-35

CRISTINE RAUBER OLIVEIRA

CORRECTED LIKELIHOOD RATIO TEST STATISTICS FOR A CLASS OF BETA REGRESSIONS WITH PARAMETRIC LINK FUNCTION

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: 17 DE FEVEREIRO DE 2020

BANCA EXAMINADORA

Prof.(º) Francisco Cribari Neto
UFPE

Prof.(ª) Patricia Leone Espinheira Ospina
UFPE

Prof.(ª) Tarciana Liberal Pereira de Araújo
UFPB

To me and my supervisors: achievement unlocked!

ACKNOWLEDGEMENTS

Undertaking this master thesis has been a life-changing experience for me and it would not have been possible without the support and guidance that I received from many people.

First and foremost, I would like to express deep gratitude to my supervisor, Professor Francisco Cribari-Neto, for his encouragement and invaluable guidance throughout this research. It was a great privilege and honor to work under his supervision. Thank you for being a patient and inspiring mentor; I am much indebted to your immense knowledge and commitment to excellence.

I would also like to gratefully acknowledge the guidance and encouragement that I received from my co-supervisor, Professor Fábio Mariano Bayer. His sage advices and excellent supervision have led me to be the successful, driven person I am today. His friendship and ongoing mentoring have been most valuable.

Francisco and Fábio, you are a source of inspiration to me on both scientific and personal grounds, and words simply cannot express my gratitude. I deeply value the efforts and dedication you have put into my education.

I would like to thank the members of my jury, Professor Patrícia Leone Espinheira Ospina and Professor Tarciana Liberal Pereira for their fruitful comments and suggestions which absolutely improved the quality of my master thesis.

I am extremely lucky to have a network of beautiful and loving friends who supported me in countless ways throughout the past two years. I will not name all of you, but I am certain that you know who you are and how important you have been to me. Thank you for all the moments we shared together. Thank you for making my journey a little less lonely. Thank you for making my studying routines much brighter, even on foggy days.

I thank the Graduate Program in Statistics of Federal University of Pernambuco (UFPE) for granting me the opportunity to work in such a compelling environment. I thank all of my professors at UFPE for their intellectual generosity, for the time they devoted to my education, and for the contributions they made to my growth as a scholar. I am thankful for the opportunities to learn and grow, and I will always look back on my years at Recife with fondness.

I am extremely grateful to my family for their love and care, for the sacrifices they made for my education, and for preparing me for the future.

Finally, I gratefully acknowledge financial support from Coordination for the Improvement of Higher Education Personnel (CAPES) – Financing Code 001.

ABSTRACT

Beta regressions are widely used for modeling random variables that assume values in the standard unit interval, $(0,1)$, such as rates, proportions, and income concentration indices. Parameter estimation is typically performed via maximum likelihood and hypothesis testing inferences on these parameters that index the model are commonly performed using the likelihood ratio test. Such a test, however, may deliver inaccurate inferences when the sample size is small. It is thus important to develop alternative testing procedures that are more accurate when the sample contains only few observations. In this master's thesis, we introduce the linear and nonlinear beta regression models with parametric mean link function, and derive two modified likelihood ratio test statistics for performing improved testing inferences in that class of models. We also obtain a score test statistic that can be used to test, in the same class of models, whether the true link function is logit. We provide simulation evidence that shows that the new tests usually outperform the standard likelihood ratio test in samples of small to moderate sizes. We also present and discuss two empirical applications.

Keywords: Beta regression. Likelihood ratio test. Link function. Parametric link function.

RESUMO

Modelos de regressão beta são amplamente utilizados para modelar variáveis aleatórias que assumem valores no intervalo unitário padrão, $(0,1)$, como taxas, proporções e índices de concentração de renda. A estimação dos parâmetros é comumente realizada através do método de máxima verossimilhança e testes de hipóteses sobre os parâmetros que indexam o modelo geralmente são realizados utilizando o teste da razão de verossimilhanças. Esse teste pode, contudo, fornecer inferências imprecisas quando o tamanho da amostra é pequeno. Portanto, é importante desenvolver testes alternativos que sejam mais precisos em situações em que a amostra contém poucas observações. Nessa dissertação, apresentamos os modelos de regressão beta linear e não linear com função de ligação paramétrica no submodelo da média, e derivamos, para uso nessa classe de modelos, duas estatísticas de teste da razão de verossimilhanças corrigidas. Obtemos também uma estatística de teste escore que, na mesma classe de modelos, pode ser usada para testar se a função de ligação do submodelo da média é logit. Fornecemos resultados numéricos que mostram que os testes modificados tipicamente apresentam desempenho superior ao do teste da razão de verossimilhanças usual em amostras de tamanho pequeno a moderado. Também apresentamos e discutimos duas aplicações em dados reais.

Palavras-chave: Função de ligação. Função de ligação paramétrica. Regressão beta. Teste da razão de verossimilhanças.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

1.1 INITIAL CONSIDERATIONS

The beta regression model introduced by Ferrari and Cribari-Neto (2004) is commonly used for modeling random variables that assume values in the unit interval, $(0,1)$. Its underlying assumption is that the dependent variable y follows the beta law, and its mean is modeled through a regression structure that involves unknown parameters, covariates, and a fixed link function. The model is also indexed by a precision parameter, which was initially taken to be constant. Extensions of the beta regression model, in its original formulation, were proposed in the literature. Ospina and Ferrari (2010) introduced the class of inflated beta distributions with support in $(0,1]$, $[0,1)$, and $[0,1]$, and Ospina and Ferrari (2012) proposed a class of regression models based on such distributions, which is known as the class of inflated beta regression models. The varying precision regression model was formally introduced by Simas, Barreto-Souza and Rocha (2010) and considered by several authors, such as Paolino (2001) and Smithson and Verkuilen (2006). Model selection strategies were developed and numerically evaluated by Bayer and Cribari-Neto (2015), Bayer and Cribari-Neto (2017), bias-corrected parameter estimation was developed by Grün, Kosmidis and Zeileis (2012) and Ospina, Cribari-Neto and Vasconcellos (2006), residuals for the model were proposed by Espinheira, Ferrari and Cribari-Neto (2007) and Espinheira, Santos and Cribari-Neto (2017), non-nested hypothesis testing strategies were developed by Cribari-Neto and Lucena (2015), bootstrap-based inferences were considered by Lima and Cribari-Neto (2020), resampling-based prediction intervals were considered by Espinheira, Ferrari and Cribari-Neto (2014), beta regression trees finite mixtures of beta regressions were considered by Grün, Kosmidis and Zeileis (2012), and time series extensions of the model were introduced by Bayer, Cintra and Cribari-Neto (2018) and Rocha and Cribari-Neto (2009) and considered by Scher *et al.* (2020). For further details, we refer readers to Cribari-Neto and Zeileis (2010).

Recently, Canterle and Bayer (2019) proposed the beta regression model with parameter-indexed link functions in the mean and dispersion submodels. Unlike the standard formulation of the model, which uses a positively-valued precision parameter, their model is indexed by a dispersion parameter that assumes values in the standard unit interval. That is, differently from the standard formulation of the beta regression model, their model is indexed by a dispersion parameter (in contrast to a precision parameter) which assumes values in $(0,1)$ (in contrast to $\mathbb{R}_+$).

Parameter estimation in beta regression analyses is usually carried out using the maximum likelihood method and hypothesis testing inferences are typically performed using the likelihood ratio test. As it is well known, such a test is based on a large sample approximation: the test critical values are obtained from the test statistic limiting null distribution. As a result, the test can be considerably size-distorted when the sample size is small. Simulation evidence reported by different authors indicates that the test tends to be liberal (oversized) in small samples, i.e., it has a tendency to overreject the null hypothesis when such a hypothesis is true. Finite sample corrections to the likelihood ratio test were considered by some authors. For instance, Ferrari and Pinheiro (2011) used approximations proposed by Skovgaard (2001) and obtained two corrected likelihood ratio test statistics for varying precision beta regressions, and Bayer and Cribari-Neto (2013) derived a Bartlett-corrected likelihood ratio test statistic (LAWLEY, 1956) for use in the fixed precision beta regression model. Improved likelihood ratio testing inference in inflated beta regressions was obtained by Pereira and Cribari-Neto (2014b).

Our first goal in this thesis is to introduce the linear and nonlinear varying precision beta regression models with parametric mean link function. We present the models and their corresponding log-likelihood function, and derive score functions, Fisher's information matrices and its inverses. The linear model differs from that of Canterle and Bayer (2019) in two ways: (i) It only uses one parametric link function and (ii) It is indexed by a precision (not dispersion) parameter that assumes values in $\mathbb{R}_+$ (not in the standard unit interval). We restrict the use of the parametric link function to the mean submodel because our chief interest involves small sample inferences, and using two parametric link functions may cause numerical instability in the parameter estimation process when the number of data points is small. Additionally, we use the same parameterization as in the beta regression formulation (FERRARI; CRIBARI-NETO, 2004), and hence the model is indexed by a positive-valued precision parameter. That way, empirical results obtained with the model can be more readily compared to those that have been already obtained using the standard beta regression model. It is also more common to index regression models by a precision or dispersion parameter that is positive-valued rather than by a parameter that assumes values in a double limited interval.

The parametric link function used in the mean submodel has two novel features that are noteworthy. The first novel feature is that it allows the link function used to model mean effects to be estimated from the data. The response and covariate values will thus determine the shape of the function that relates the mean of the dependent variable to the linear predictor, including the degree of asymmetry (if any) of such a relationship. There is typically no or very little information on the link function that relates the response mean to the corresponding linear predictor and an incorrect choice may render the model misspecified. For instance, the use of a symmetric link function when the true relationship between the mean and the predictor shows asymmetry may impact statistical inferences and predictions. A way to circumvent that problem is to define a data-driven link function that can be symmetric or asymmetric. This is the approach we shall pursue. The second novel feature of the parameter-indexed link function we use is that it includes the logit link function, which is most commonly used function in applications, as a particular case. It is then possible to test whether the link function is logit by performing a test on the parameter that indexes the parametric link function. We derive a score test statistic that can be used to that end. We note that the use of parametric link functions in generalized linear models (MCCULLAGH; NELDER, 1989) was initially investigated by Pregibon (1980) and further advanced by Czado (1992) and Czado (1997). Colosimo, Chalita and Demétrio (2000) derived score test statistics using a data-driven link function for grouped survival applications. More recently, Dehbi, Cortina-Borja and Geraci (2016) modeled student performance assessment using quantile regression with parametric link families.

Our second goal relates to testing inferences. In order to achieve accurate testing inferences in small samples, we derive two improved likelihood ratio test statistics using general results obtained by Skovgaard (2001). They can be used to test restrictions on the parameters that index the two submodels of the varying precision beta regression model with parametric link function. We also derive a score test statistic to test that the mean link function is logit in our models. Our Monte Carlo results evidence shows that the corrected tests typically outperform the standard likelihood ratio test when the number of observations is small, i.e., they tend to be less size-distorted. For example, in one of the configurations and based on 30 data points, the size distortion (difference between exact and nominal sizes) of the likelihood ratio test at the 10% significance level exceeded 15% (i.e., its estimated size exceeded 25%) whereas the size distortions of the two tests we develop in this paper were -1% and 1.5%. These numerical results show the usefulness of the two improved tests.

In addition to Monte Carlo evidence, we present and discuss two empirical applications. In both of them, superior model fit is achieved by letting the relationship between the mean response and the corresponding linear predictor be data-driven. Our second application uses data on the data on the prevalence of religious disbelievers and average intelligence in 124 countries. Using such data, Cribari-Neto and Souza (2013) computed a measure of impact of intelligence on atheism. The estimated impact was displayed as a function of average intelligence and showed a bell-curve relationship: the impact strengthens as average intelligence grows, peaks, and then weakens. By using a parametric mean link function we were able to uncover an existing asymmetry regarding the rates at which the impact strengthens and weakens prior and after peaking, respectively. In particular, we show that the impact weakens at a slower rate.

1.2 THESIS ORGANIZATION

The thesis is structured as follows. Chapter 2 presents the linear and nonlinear varying precision beta regression models with parametric mean link function. In Chapter 3 we derive two adjusted likelihood ratio test statistics, and for the particular case of testing whether the link function is logit we also derive the score test statistic for use with our models. Chapter 4 contains Monte Carlo simulation evidence on the small sample behavior of the modified tests and of the likelihood ratio test. We also provide evidence on the small sample behavior of a score test. In Chapter 5 we present and discuss two empirical applications. Finally, some concluding remarks are offered in Chapter 6. The Appendix contains the computer code used in the first empirical application.

1.3 COMPUTING PLATFORMS

The Monte Carlo simulations and empirical applications were carried out using the Ox matrix programming language (version 8.02) for the Linux operating system; see Doornik (2009) for details. All figures presented in this thesis were produced using the R statistical computing environment (R Core Team, 2018). This master's thesis was typeset using LaTeX.

2 BETA REGRESSION MODELS WITH PARAMETRIC MEAN LINK FUNCTION

The beta law is commonly used with data that are restricted to the standard unit interval. Its density can assume a wide variety of shapes as the values of the parameters that index the law vary. In particular, it can be asymmetric (to the right or to the left), symmetric, J-shaped, inverted J-shaped, U-shaped, and uniform. An alternative parameterization of the beta law was considered by Ferrari and Cribari-Neto (2004) in which the beta density is indexed by a location (μ) and a precision (ϕ) parameter. The authors wrote the beta density as

$$b(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \qquad 0 < y < 1, \tag{2.1}$$

where $0 < \mu < 1$, $\phi > 0$, and $\Gamma(\cdot)$ is the gamma function. The distribution mean is μ and the variance is $[\mu(1-\mu)]/(1+\phi)$. We shall denote the beta law based on such a parameterization as $\mathcal{B}(\mu,\phi)$.

2.1 LINEAR VARYING PRECISION BETA MODEL WITH PARAMETRIC MEAN LINK FUNCTION

Let $y_1,\ldots,y_n$ be a set of independent random variables such that $y_t \sim \mathcal{B}(\mu_t,\phi_t)$, $t = 1,\ldots,n$. The linear varying precision beta regression model with parametric mean link function is given by

$$g(\mu_t,\lambda) = \sum_{i=1}^{p} x_{ti}\beta_i = \eta_{1t}, \tag{2.2}$$

$$h(\phi_t) = \sum_{j=1}^{q} z_{tj}\gamma_j = \eta_{2t}, \tag{2.3}$$

where $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)^\top \in \mathbb{R}^p$ and $\boldsymbol{\gamma} = (\gamma_1,\ldots,\gamma_q)^\top \in \mathbb{R}^q$ are unknown parameters vectors ($p + q + 1 = k < n$), $\boldsymbol{\eta_1} = (\eta_{11},\ldots,\eta_{1n})^\top$ and $\boldsymbol{\eta_2} = (\eta_{21},\ldots,\eta_{2n})^\top$ being the mean and precision linear predictor vectors, respectively. Here, $g(\cdot,\cdot)$ and $h(\cdot)$ are link functions that are strictly monotonic in the first argument and twice differentiable (the former in both arguments) such that $g : (0,1)\times(0,\infty) \mapsto \mathbb{R}$ and $h : (0,\infty) \mapsto \mathbb{R}$. The regression structures for μ_t and ϕ_t can then be expressed as

$$\mu_t = g^{-1}(\eta_{1t},\lambda), \tag{2.4}$$

$$\phi_t = h^{-1}(\eta_{2t}). \tag{2.5}$$

Common choices for the precision link function are log and square root, i.e., $h(\phi_t) = \log(\phi_t)$ and $h(\phi_t) = \sqrt{\phi_t}$, respectively.

This model differs from the beta regression that is commonly used in the literature because we consider a mean link function (g) that includes a parameter to be estimated from the data. In particular, as in Canterle and Bayer (2019), we consider the asymmetric Aranda-Ordaz (AO) link function which is given by (ARANDA-ORDAZ, 1981)

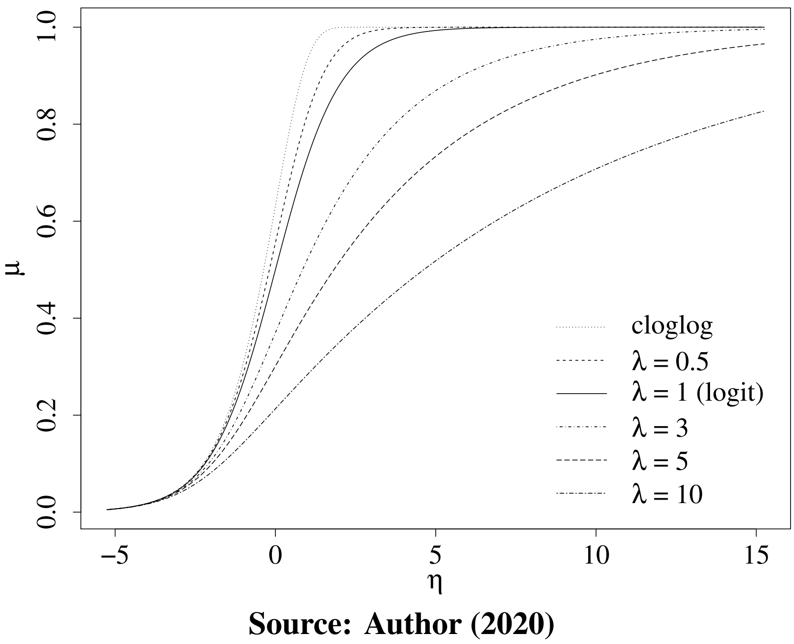
$$g(\mu_t, \lambda) = \log\left(\frac{(1-\mu_t)^{-\lambda} - 1}{\lambda}\right), \tag{2.6}$$

where $\lambda > 0$. The inverse mean link function is

$$g^{-1}(\eta_{1t}, \lambda) = 1 - (1 + \lambda e^{\eta_{1t}})^{-1/\lambda}. \tag{2.7}$$

A noteworthy advantage of the AO family of link functions is that it includes two well known link functions as particular cases: (i) the logit link follows from setting $\lambda = 1$, and (ii) the complementary loglog (cloglog) link is obtained by letting $\lambda \to 0$. Additionally, the family includes asymmetric link functions that can capture existing asymmetries in the relationship between the mean response and the corresponding linear predictor. Figure 1 shows the relationship between η and μ for different values of λ. Notice that the convergence of μ to one as η increases happens at progressively slower paces as λ increases. The estimated parametric mean link accommodates the convergence rate and the corresponding degree of asymmetry that are best suited to the data at hand.

Figure 1 – AO link function for different values of λ.



Source: Author (2020)

We shall now develop statistical inference for the linear varying precision beta regression model with parametric mean link function. In particular, we shall present the model log-likelihood function and obtain the corresponding score function, and Fisher's information matrix. The inverse information matrix will also be presented in closed-form.

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \lambda)^\top$ be the beta regression parameter vector. The log-likelihood function of $\boldsymbol{\theta}$ based on n independent observations is

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^{n} \ell_t(\mu_t, \phi_t), \tag{2.8}$$

where

$$\ell_t(\mu_t, \phi_t) = \log\Gamma(\phi_t) - \log\Gamma(\mu_t\phi_t) - \log\Gamma((1-\mu_t)\phi_t) + (\mu_t\phi_t - 1)\log y_t + [(1-\mu_t)\phi_t - 1]\log(1-y_t), \tag{2.9}$$

with μ_t and ϕ_t being given in (2.4) and (2.5), respectively.

By letting $y_t^* = \log(y_t/(1-y_t))$ and $y_t^\dagger = \log(1-y_t)$, it is possible to write the above log-likelihood function as

$$\ell_t(\mu_t, \phi_t) = \log\Gamma(\phi_t) - \log\Gamma(\mu_t\phi_t) - \log\Gamma((1-\mu_t)\phi_t) + (\mu_t\phi_t - 1)y_t^* + (\phi_t - 2)y_t^\dagger. \tag{2.10}$$

We note that the moments of y^* and $y^\dagger$ are easily obtained. It can be shown that $\mu_t^* = \mathbb{E}(y_t^*) = \psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t)$, $\mu_t^\dagger = \mathbb{E}(y_t^\dagger) = \psi((1-\mu_t)\phi_t) - \psi(\phi_t)$, $\upsilon_t^* = \mathrm{Var}(y_t^*) = \psi'(\mu_t\phi_t) + \psi'((1-\mu_t)\phi_t)$, $\upsilon_t^\dagger = \mathrm{Var}(y_t^\dagger) = \psi'((1-\mu_t)\phi_t) - \psi'(\phi_t)$, and $c_t^{*\dagger} = \mathrm{cov}(y_t^*, y_t^\dagger) = -\psi'((1-\mu_t)\phi_t)$, where ψ and ψ' denote the digamma and trigamma functions, respectively.

The model log-likelihood function can be expressed in matrix form as (FERRARI; PINHEIRO, 2011)

$$\ell(\boldsymbol{\theta}) = \left[(\boldsymbol{y}^* - \boldsymbol{\mu}^*)^\top(\Phi M - I_n) + (\boldsymbol{y}^\dagger - \boldsymbol{\mu}^\dagger)^\top(\Phi - 2I_n) + \boldsymbol{b}^\top\right]\boldsymbol{\iota}, \tag{2.11}$$

where $\boldsymbol{y}^* = (y_1^*, \ldots, y_n^*)^\top$, $\boldsymbol{\mu}^* = (\mu_1^*, \ldots, \mu_n^*)^\top$, $\boldsymbol{y}^\dagger = (y_1^\dagger, \ldots, y_n^\dagger)^\top$, $\boldsymbol{\mu}^\dagger = (\mu_1^\dagger, \ldots, \mu_n^\dagger)^\top$, $\Phi = \mathrm{diag}(\phi_1, \ldots, \phi_n)$, $M = \mathrm{diag}(\mu_1, \ldots, \mu_n)$, I_n is the $n \times n$ identity matrix, $\boldsymbol{\iota}$ denotes a n-dimensional vector of 1's, and $\boldsymbol{b} = (b_1, \ldots, b_n)^\top$, with $b_t = \log\Gamma(\phi_t) - \log\Gamma(\mu_t\phi_t) - \log\Gamma((1-\mu_t)\phi_t) + \mu_t^*(\mu_t\phi_t - 1) + \mu_t^\dagger(\phi_t - 2)$. Maximum likelihood estimates can be obtained by numerically maximizing (2.11) with respect to $\boldsymbol{\theta}$ using a Newton or quasi-Newton algorithm. Notice that λ is estimated jointly with $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

We shall now obtain the score function by differentiating the log-likelihood function in (2.8) with respect to each component of $\boldsymbol{\theta}$. For $i = 1, \ldots, p$ and $j = 1, \ldots, q$, we obtain

$$U_{\beta_i}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_i} = \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \beta_i} = \sum_{t=1}^{n} \left\{ \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{\partial \mu_t}{\partial \eta_{1t}} \frac{\partial \eta_{1t}}{\partial \beta_i} \right\}, \tag{2.12}$$

$$U_{\gamma_j}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma_j} = \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \gamma_j} = \sum_{t=1}^{n} \left\{ \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} \frac{\partial \phi_t}{\partial \eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_j} \right\}, \tag{2.13}$$

$$U_{\lambda}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \lambda} = \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \lambda} = \sum_{t=1}^{n} \left\{ \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{\partial \mu_t}{\partial \lambda} \right\}, \tag{2.14}$$

where $\dfrac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} = \phi_t(y_t^* - \mu_t^*)$, $\dfrac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} = \mu_t(y_t^* - \mu_t^*) + (y_t^\dagger - \mu_t^\dagger)$, $\dfrac{\partial \mu_t}{\partial \eta_{1t}} = \dfrac{1}{g'(\mu_t, \lambda)}$, $\dfrac{\partial \phi_t}{\partial \eta_{2t}} = \dfrac{1}{h'(\phi_t)}$, $\dfrac{\partial \mu_t}{\partial \lambda} = \rho_t = \dfrac{1}{\lambda}\left[\dfrac{1}{e^{-\eta_{1t}} + \lambda} - \dfrac{\log(1 + \lambda e^{\eta_{1t}})}{\lambda}\right](1 + \lambda e^{\eta_{1t}})^{-1/\lambda}$, $\dfrac{\partial \eta_{1t}}{\partial \beta_i} = x_{ti}$, and $\dfrac{\partial \eta_{2t}}{\partial \gamma_j} = z_{tj}$. Therefore,

$$\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \beta_i} = \sum_{t=1}^{n} \left\{ \phi_t(y_t^* - \mu_t^*) \frac{1}{g'(\mu_t, \lambda)} x_{ti} \right\}, \tag{2.15}$$

$$\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \gamma_j} = \sum_{t=1}^{n} \left\{ \left[\mu_t(y_t^* - \mu_t^*) + (y_t^\dagger - \mu_t^\dagger) \right] \frac{1}{h'(\phi_t)} z_{tj} \right\}, \tag{2.16}$$

$$\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \lambda} = \sum_{t=1}^{n} \left\{ \phi_t(y_t^* - \mu_t^*) \rho_t \right\}, \tag{2.17}$$

where $g'(\mu_t, \lambda) = \dfrac{\partial g(\mu_t, \lambda)}{\partial \mu_t} = \dfrac{\lambda(1 - \mu_t)^{-(\lambda+1)}}{(1 - \mu_t)^{-\lambda} - 1}$ and $h'(\phi_t) = \dfrac{\partial h(\phi_t)}{\partial \phi_t}$.

The score function is $\boldsymbol{U} \equiv \boldsymbol{U}(\boldsymbol{\theta}) = \left(\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta})^\top, \boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta})^\top, \boldsymbol{U}_{\lambda}(\boldsymbol{\theta})\right)^\top$. It can be expressed in matrix form as follows

$$\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = X^\top \Phi T(\boldsymbol{y}^* - \boldsymbol{\mu}^*), \tag{2.18}$$

$$\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}) = Z^\top H \left[M(\boldsymbol{y}^* - \boldsymbol{\mu}^*) + (\boldsymbol{y}^\dagger - \boldsymbol{\mu}^\dagger) \right], \tag{2.19}$$

$$\boldsymbol{U}_{\lambda}(\boldsymbol{\theta}) = \boldsymbol{\rho}^\top \Phi(\boldsymbol{y}^* - \boldsymbol{\mu}^*), \tag{2.20}$$

where X is the $n \times p$ matrix of mean covariates whose the tth row is x_t, Z is the $n \times q$ matrix of precision covariates whose tth row is z_t, $T = \mathrm{diag}\left(\dfrac{1}{g'(\mu_1, \lambda)}, \ldots, \dfrac{1}{g'(\mu_n, \lambda)}\right)$, $H = \mathrm{diag}\left(\dfrac{1}{h'(\phi_1)}, \ldots, \dfrac{1}{h'(\phi_n)}\right)$, and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_n)^\top$.

The maximum likelihood estimators (MLE) of the parameters that index the beta regression model with parametric mean link function solve

$$\begin{cases} \boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \mathbf{0} \\ \boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}) &= \mathbf{0} \\ \boldsymbol{U}_{\lambda}(\boldsymbol{\theta}) &= \mathbf{0}, \end{cases} \tag{2.21}$$

where $\mathbf{0}$ denotes a column-vector of zeros of the appropriate dimension. Since the above system of equations does not have a closed-form solution, maximum likelihood estimation can be performed by numerically maximizing the log-likelihood function using a nonlinear optimization algorithm.

The Hessian matrix is obtained by taking the second order derivatives of (2.8) with respect to the components of $\boldsymbol{\theta}$. For $i,r = 1,\ldots,p$, we have

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\beta_i\partial\beta_r} &= \sum_{t=1}^{n}\left\{\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\beta_i\partial\beta_r}\right\},\\
&= \sum_{t=1}^{n}\left\{\frac{\partial}{\partial\beta_r}\left[\frac{\partial \ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\beta_i}\right]\right\},\\
&= \sum_{t=1}^{n}\left\{\frac{\partial}{\partial\mu_t}\left[\frac{\partial \ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial\mu_t}{\partial\eta_{1t}}\right]\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\beta_r}\frac{\partial\eta_{1t}}{\partial\beta_i}\right\},\\
&= \sum_{t=1}^{n}\left\{\left[\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\mu_t^2}\frac{\partial\mu_t}{\partial\eta_{1t}}+\frac{\partial \ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial^2\mu_t}{\partial\eta_{1t}\partial\mu_t}\right]\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\beta_r}\frac{\partial\eta_{1t}}{\partial\beta_i}\right\},
\end{aligned}
$$

where $\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\mu_t^2} = -\phi_t^2\upsilon_t^*$ and $\frac{\partial^2\mu_t}{\partial\eta_{1t}\partial\mu_t} = -\frac{g''(\mu_t,\lambda)}{g'(\mu_t,\lambda)^2}$, with $g''(\mu_t,\lambda) = \frac{\partial^2 g(\mu_t,\lambda)}{\partial\mu_t^2} = \frac{\lambda-(1-\mu_t)^{\lambda}\lambda(1+\lambda)}{(\mu_t-1)^2[(1-\mu_t)^{\lambda}-1]^2}$. Therefore,

$$
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\beta_i\partial\beta_r} = \sum_{t=1}^{n}\left\{\left[-\phi_t^2\upsilon_t^*\left(\frac{1}{g'(\mu_t,\lambda)}\right)+\phi_t(y_t^*-\mu_t^*)\left(-\frac{g''(\mu_t,\lambda)}{g'(\mu_t,\lambda)^2}\right)\right]\left(\frac{1}{g'(\mu_t,\lambda)}\right)x_{tr}x_{ti}\right\}.
$$

Additionally, for $i = 1,\ldots,p$ and $j = 1,\ldots,q$, we obtain

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\beta_i\partial\gamma_j} &= \sum_{t=1}^{n}\left\{\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\beta_i\partial\gamma_j}\right\},\\
&= \sum_{t=1}^{n}\left\{\frac{\partial}{\partial\gamma_j}\left[\frac{\partial \ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\beta_i}\right]\right\},\\
&= \sum_{t=1}^{n}\left\{\left[\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\mu_t\partial\gamma_j}\right]\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\beta_i}\right\},\\
&= \sum_{t=1}^{n}\left\{\left[\frac{\partial}{\partial\phi_t}\left(\frac{\partial \ell_t(\mu_t,\phi_t)}{\partial\mu_t}\right)\frac{\partial\phi_t}{\partial\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\gamma_j}\right]\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\beta_i}\right\},\\
&= \sum_{t=1}^{n}\left\{\left[\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\mu_t\partial\phi_t}\frac{\partial\phi_t}{\partial\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\gamma_j}\right]\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\beta_i}\right\},
\end{aligned}
$$

where $\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\mu_t\partial\phi_t} = (y_t^*-\mu_t^*)-\phi_t(\mu_t\upsilon_t^*+c_t^{*\dagger})$. Hence,

$$
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\beta_i\partial\gamma_j} = \sum_{t=1}^{n}\left\{\left[(y_t^*-\mu_t^*)-\phi_t(\mu_t\upsilon_t^*+c_t^{*\dagger})\right]\left(\frac{1}{h'(\phi_t)}\right)z_{tj}\left(\frac{1}{g'(\mu_t,\lambda)}\right)x_{ti}\right\}.
$$

Also,

$$
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\beta_i\partial\lambda} = \sum_{t=1}^{n}\left\{\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial\beta_i\partial\lambda}\right\},
$$

$$= \sum_{t=1}^{n} \left\{ \frac{\partial}{\partial \lambda} \left[\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{\partial \mu_t}{\partial \eta_{1t}} \frac{\partial \eta_{1t}}{\partial \beta_i} \right] \right\},$$

$$= \sum_{t=1}^{n} \left\{ \left[\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \mu_t^2} \frac{\partial \mu_t}{\partial \lambda} \frac{\partial \mu_t}{\partial \eta_{1t}} + \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{\partial}{\partial \lambda} \left(\frac{\partial \mu_t}{\partial \eta_{1t}} \right) \right] \frac{\partial \eta_{1t}}{\partial \beta_i} \right\},$$

where $\frac{\partial}{\partial \lambda} \left(\frac{\partial \mu_t}{\partial \eta_{1t}} \right) = w_t = \frac{e^{\eta_{1t}}(1+\lambda e^{\eta_{1t}})^{-2-\frac{1}{\lambda}}}{\lambda^2} \left[\left(-\lambda e^{\eta_{1t}}(1+\lambda) + (1+\lambda e^{\eta_{1t}}) \log(1+\lambda e^{\eta_{1t}} \right) \right]$. Therefore,

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_i \partial \lambda} = \sum_{t=1}^{n} \left\{ \left[-\phi_t^2 \upsilon_t^* \rho_t \left(\frac{1}{g'(\mu_t, \lambda)} \right) + \phi_t (y_t^* - \mu_t^*) \frac{\partial}{\partial \lambda} \left(\frac{1}{g'(\mu_t, \lambda)} \right) \right] x_{ti} \right\}.$$

Similarly, for $j, g = 1, \ldots, q$, we have

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_g} = \sum_{t=1}^{n} \left\{ \frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \gamma_j \partial \gamma_g} \right\},$$

$$= \sum_{t=1}^{n} \left\{ \frac{\partial}{\partial \gamma_g} \left[\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} \frac{\partial \phi_t}{\partial \eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_j} \right] \right\},$$

$$= \sum_{t=1}^{n} \left\{ \frac{\partial}{\partial \phi_t} \left[\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} \frac{\partial \phi_t}{\partial \eta_{2t}} \right] \frac{\partial \phi_t}{\partial \eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_g} \frac{\partial \eta_{2t}}{\partial \gamma_j} \right\},$$

$$= \sum_{t=1}^{n} \left\{ \left[\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \phi_t^2} \frac{\partial \phi_t}{\partial \eta_{2t}} + \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} \frac{\partial^2 \phi_t}{\partial \eta_{2t} \partial \phi_t} \right] \frac{\partial \phi_t}{\partial \eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_g} \frac{\partial \eta_{2t}}{\partial \gamma_j} \right\},$$

where $\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \phi_t^2} = -\mu_t^2 v_t^* - 2\mu_t c_t^{*\dagger} - v^\dagger$ and $\frac{\partial^2 \phi_t}{\partial \eta_{2t} \partial \phi_t} = -\frac{h''(\phi_t)}{h'(\phi_t)^2}$, with $h''(\phi_t) = \frac{\partial^2 h(\phi_t)}{\partial \phi_t^2}$. Hence,

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_j \partial \gamma_g} = \sum_{t=1}^{n} \left\{ \left[\left(-\mu_t^2 v_t^* - 2\mu_t c_t^{*\dagger} - v^\dagger \right) \left(\frac{1}{h'(\phi_t)} \right) + \left(\mu_t (y_t^* - \mu_t^*) + (y_t^\dagger - \mu_t^\dagger) \right) \left(-\frac{h''(\phi_t)}{h'(\phi_t)^2} \right) \right] \right.$$
$$\left. \times \left(\frac{1}{h'(\phi_t)} \right) z_{tg} z_{tj} \right\}.$$

Also,

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_j \partial \lambda} = \sum_{t=1}^{n} \left\{ \frac{\partial^2 \ell(\mu_t, \phi_t)}{\partial \gamma_j \partial \lambda} \right\},$$

$$= \sum_{t=1}^{n} \left\{ \frac{\partial}{\partial \lambda} \left[\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} \frac{\partial \phi_t}{\partial \eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_j} \right] \right\},$$

$$= \sum_{t=1}^{n} \left\{ \frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \phi_t \partial \mu_t} \frac{\partial \mu_t}{\partial \lambda} \frac{\partial \phi_t}{\partial \eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_j} \right\},$$

where $\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \phi_t \partial \mu_t} = \frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \mu_t \partial \phi_t}$. Therefore,

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_j \partial \lambda} = \sum_{t=1}^{n} \left\{ \left[(y_t^* - \mu_t^*) - \phi_t (\mu_t v_t^* + c_t^{*\dagger}) \right] \rho_t \left(\frac{1}{h'(\phi_t)} \right) z_{tj} \right\}.$$

Finally,

$$\begin{aligned}\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \lambda^2} &= \sum_{t=1}^{n}\left\{\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial \lambda^2}\right\},\\ &= \sum_{t=1}^{n}\left\{\frac{\partial}{\partial \lambda}\left(\frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \mu_t}\frac{\partial \mu_t}{\partial \lambda}\right)\right\},\\ &= \sum_{t=1}^{n}\left\{\frac{\partial^2 \ell_t(\mu_t,\phi_t)}{\partial \mu_t^2}\frac{\partial \mu_t}{\partial \lambda}\frac{\partial \mu_t}{\partial \lambda} + \frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \mu_t}\frac{\partial^2 \mu}{\partial \lambda^2}\right\},\end{aligned}$$

where

$$\frac{\partial^2 \mu_t}{\partial \lambda^2} = \varrho_t = \frac{(1+\lambda e^{\eta_{1t}})^{-\frac{1}{\lambda}-2}}{\lambda^4}\big[(1+\lambda e^{\eta_{1t}})\log(1+\lambda e^{\eta_{1t}})(2\lambda(1+e^{\eta_{1t}}(1+\lambda)) - (1+\lambda e^{\eta_{1t}})$$
$$\times \log(1+\lambda e^{\eta_{1t}})) - \lambda^2 e^{\eta_{1t}}((3\lambda+1)e^{\eta_{1t}}+2)\big].$$

Hence,

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \lambda^2} = \sum_{t=1}^{n}\left\{-\phi_t^2 \upsilon_t^* \rho_t \rho_t + \phi_t (y_t^* - \mu_t^*)\frac{\partial^2 \mu}{\partial \lambda^2}\right\}.$$

The negative Hessian is the observed information matrix. It can be written in matrix form as

$$J \equiv J(\boldsymbol{\theta}) = \begin{pmatrix} J_{(\boldsymbol{\beta},\boldsymbol{\beta})} & J_{(\boldsymbol{\beta},\boldsymbol{\gamma})} & J_{(\boldsymbol{\beta},\lambda)} \\ J_{(\boldsymbol{\gamma},\boldsymbol{\beta})} & J_{(\boldsymbol{\gamma},\boldsymbol{\gamma})} & J_{(\boldsymbol{\gamma},\lambda)} \\ J_{(\lambda,\boldsymbol{\beta})} & J_{(\lambda,\boldsymbol{\gamma})} & J_{(\lambda,\lambda)} \end{pmatrix},$$

where $J_{(\boldsymbol{\beta},\boldsymbol{\beta})} = X^\top[\Phi T V^* + S T^2 (Y^* - M^*)]T\Phi X$, $J_{(\boldsymbol{\beta},\boldsymbol{\gamma})} = J_{(\boldsymbol{\gamma},\boldsymbol{\beta})}^\top = -X^\top[(Y^* - M^*) - \Phi(MV^* + C)]THZ$, $J_{(\boldsymbol{\beta},\lambda)} = J_{(\lambda,\boldsymbol{\beta})}^\top = X^\top[\Phi^2 V^* T\boldsymbol{\rho} - \Phi(Y^* - M^*)\boldsymbol{w}]$, $J_{(\boldsymbol{\gamma},\boldsymbol{\gamma})} = Z^\top\{H(M^2V^* + 2MC + V^\dagger) + [M(Y^* - M^*) + (Y^\dagger - M^\dagger)]H^2 Q\}HZ$, $J_{(\boldsymbol{\gamma},\lambda)} = J_{(\lambda,\boldsymbol{\gamma})}^\top = -Z^\top[(Y^* - M^*) - \Phi(MV^* + C)]H\boldsymbol{\rho}$, and $J_{(\lambda,\lambda)} = [\Phi^2 V^* \boldsymbol{\rho}^2 - \Phi(Y^* - M^*)\boldsymbol{\varrho}]^\top \boldsymbol{\imath}$. Here, $Y^* = \mathrm{diag}(y_1^*,\ldots,y_n^*)$, $Y^\dagger = \mathrm{diag}(y_1^\dagger,\ldots,y_n^\dagger)$, $M^* = \mathrm{diag}(\mu_1^*,\ldots,\mu_n^*)$, $M^\dagger = \mathrm{diag}(\mu_1^\dagger,\ldots,\mu_n^\dagger)$, $V^* = \mathrm{diag}(\upsilon_1^*,\ldots,\upsilon_n^*)$, $V^\dagger = \mathrm{diag}(\upsilon_1^\dagger,\ldots,\upsilon_n^\dagger)$, $C = \mathrm{diag}(c_1^{*\dagger}, \ldots, c_n^{*\dagger})$, $S = \mathrm{diag}(g''(\mu_1,\lambda),\ldots,g''(\mu_n,\lambda))$, $Q = \mathrm{diag}(h''(\phi_1),\ldots,h''(\phi_n))$, $\boldsymbol{w} = (w_1,\ldots,w_n)^\top$, and $\boldsymbol{\varrho} = (\varrho_1,\ldots,\varrho_n)^\top$.

Since $\mathbb{E}\left(\frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \mu_t}\right) = \mathbb{E}\left(\frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \phi_t}\right) = 0$, Fisher's information matrix is given by

$$K \equiv K(\boldsymbol{\theta}) = \begin{pmatrix} K_{(\boldsymbol{\beta},\boldsymbol{\beta})} & K_{(\boldsymbol{\beta},\boldsymbol{\gamma})} & K_{(\boldsymbol{\beta},\lambda)} \\ K_{(\boldsymbol{\gamma},\boldsymbol{\beta})} & K_{(\boldsymbol{\gamma},\boldsymbol{\gamma})} & K_{(\boldsymbol{\gamma},\lambda)} \\ K_{(\lambda,\boldsymbol{\beta})} & K_{(\lambda,\boldsymbol{\gamma})} & K_{(\lambda,\lambda)} \end{pmatrix}, \tag{2.22}$$

where $K_{(\boldsymbol{\beta},\boldsymbol{\beta})} = X^\top \Phi^2 V^* T^2 X$, $K_{(\boldsymbol{\beta},\boldsymbol{\gamma})} = K_{(\boldsymbol{\gamma},\boldsymbol{\beta})}^\top = X^\top \Phi(MV^* + C)THZ$, $K_{(\boldsymbol{\beta},\lambda)} = K_{(\lambda,\boldsymbol{\beta})}^\top = X^\top \Phi^2 V^* T\boldsymbol{\rho}$, $K_{(\boldsymbol{\gamma},\boldsymbol{\gamma})} = Z^\top H(M^2V^* + 2MC + V^\dagger)HZ$, $K_{(\boldsymbol{\gamma},\lambda)} = K_{(\lambda,\boldsymbol{\gamma})}^\top = Z^\top \Phi(MV^* + C)H\boldsymbol{\rho}$, and $K_{(\lambda,\lambda)} = \boldsymbol{\rho}^\top \Phi^2 V^* \boldsymbol{\rho}$.

In large samples and under the usual regularity conditions for maximum likelihood estimation, we have

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \\ \hat{\lambda} \end{pmatrix} \sim \mathcal{N}_k\left(\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \\ \lambda \end{pmatrix}, K^{-1} \right),$$

approximately, where $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}$, and $\hat{\lambda}$ are the MLEs of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and λ, respectively, $\mathcal{N}_k$ denoting the k-variate normal distribution. In what follows, we shall use a result on inverses of partitioned matrices given by Rao (1973, p. 33) to obtain a closed-form expression for K^{-1}. Such a matrix is useful, for instance, for computing asymptotic standard errors for the point estimates.

Consider the symmetric matrix given by

$$I \equiv I(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \begin{pmatrix} I_{(\boldsymbol{\beta},\boldsymbol{\beta})} & I_{(\boldsymbol{\beta},\boldsymbol{\gamma})} \\ I_{(\boldsymbol{\gamma},\boldsymbol{\beta})} & I_{(\boldsymbol{\gamma},\boldsymbol{\gamma})} \end{pmatrix},$$

where $I_{(\boldsymbol{\beta},\boldsymbol{\beta})} = K_{(\boldsymbol{\beta},\boldsymbol{\beta})}$, $I_{(\boldsymbol{\beta},\boldsymbol{\gamma})} = K_{(\boldsymbol{\beta},\boldsymbol{\gamma})}$, $I_{(\boldsymbol{\gamma},\boldsymbol{\beta})} = K_{(\boldsymbol{\gamma},\boldsymbol{\beta})}$, and $I_{(\boldsymbol{\gamma},\boldsymbol{\gamma})} = K_{(\boldsymbol{\gamma},\boldsymbol{\gamma})}$. We denote its inverse as

$$I^{-1} \equiv I^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \begin{pmatrix} I^{(\boldsymbol{\beta},\boldsymbol{\beta})} & I^{(\boldsymbol{\beta},\boldsymbol{\gamma})} \\ I^{(\boldsymbol{\gamma},\boldsymbol{\beta})} & I^{(\boldsymbol{\gamma},\boldsymbol{\gamma})} \end{pmatrix}.$$

It can be shown that

$$I^{(\boldsymbol{\beta},\boldsymbol{\beta})} = (X^\top \Phi^2 V^* T^2 X)^{-1} \left[I_p + \frac{X^\top \Phi (MV^* + C) THZZ^\top H^\top T^\top (MV^* + C)^\top \Phi^\top X (X^\top \Phi^2 V^* T^2 X)^{-1}}{\omega} \right],$$

with $\omega = Z^\top H(M^2V^* + 2MC + V^\dagger)HZ - Z^\top H^\top T^\top (MV^* + C)^\top \Phi^\top X (X^\top \Phi^2 V^* T^2 X)^{-1} X^\top \Phi (MV^* + C)THZ$, I_p denoting the $p \times p$ identity matrix. Additionally,

$$I^{(\boldsymbol{\beta},\boldsymbol{\gamma})} = (I^{(\boldsymbol{\gamma},\boldsymbol{\beta})})^\top = -\frac{(X^\top \Phi^2 V^* T^2 X)^{-1} X^\top \Phi (MV^* + C)THZ}{\omega} \quad \text{and} \quad I^{(\boldsymbol{\gamma},\boldsymbol{\gamma})} = \omega^{-1}.$$

We obtain

$$K^{-1} \equiv K^{-1}(\boldsymbol{\theta}) = \begin{pmatrix} K^{(\boldsymbol{\beta},\boldsymbol{\beta})} & K^{(\boldsymbol{\beta},\boldsymbol{\gamma})} & K^{(\boldsymbol{\beta},\lambda)} \\ K^{(\boldsymbol{\gamma},\boldsymbol{\beta})} & K^{(\boldsymbol{\gamma},\boldsymbol{\gamma})} & K^{(\boldsymbol{\gamma},\lambda)} \\ K^{(\lambda,\boldsymbol{\beta})} & K^{(\lambda,\boldsymbol{\gamma})} & K^{(\lambda,\lambda)} \end{pmatrix}, \tag{2.23}$$

where

$$K^{(\boldsymbol{\beta},\boldsymbol{\beta})} = I^{(\boldsymbol{\beta},\boldsymbol{\beta})} + \Omega[I^{(\boldsymbol{\beta},\boldsymbol{\beta})} X^\top \Phi^2 V^* T \boldsymbol{\rho} + I^{(\boldsymbol{\beta},\boldsymbol{\gamma})} Z^\top \Phi (MV^* + C) H \boldsymbol{\rho}]$$
$$\times [\boldsymbol{\rho}^\top T^\top (V^*)^\top (\Phi^2)^\top X I^{(\boldsymbol{\beta},\boldsymbol{\beta})} + \boldsymbol{\rho}^\top H^\top (MV^* + C)^\top \Phi^\top Z I^{(\boldsymbol{\gamma},\boldsymbol{\beta})}],$$
$$K^{(\boldsymbol{\gamma},\boldsymbol{\gamma})} = I^{(\boldsymbol{\gamma},\boldsymbol{\gamma})} + \Omega[I^{(\boldsymbol{\gamma},\boldsymbol{\beta})} X^\top \Phi^2 V^* T \boldsymbol{\rho} + I^{(\boldsymbol{\gamma},\boldsymbol{\gamma})} Z^\top \Phi (MV^* + C) H \boldsymbol{\rho}]$$

$$
\begin{aligned}
&\times[\boldsymbol{\rho}^\top T^\top (V^*)^\top (\Phi^2)^\top X I^{(\boldsymbol{\beta},\boldsymbol{\gamma})} + \boldsymbol{\rho}^\top H^\top (MV^*+C)^\top \Phi^\top Z I^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}],\\
K^{(\boldsymbol{\beta},\boldsymbol{\gamma})} &= (K^{(\boldsymbol{\gamma},\boldsymbol{\beta})})^\top = I^{(\boldsymbol{\beta},\boldsymbol{\gamma})} + \Omega[I^{(\boldsymbol{\beta},\boldsymbol{\beta})} X^\top \Phi^2 V^* T\boldsymbol{\rho} + I^{(\boldsymbol{\beta},\boldsymbol{\gamma})} Z^\top \Phi(MV^*+C)H\boldsymbol{\rho}]\\
&\times[\boldsymbol{\rho}^\top T^\top (V^*)^\top (\Phi^2)^\top X I^{(\boldsymbol{\beta},\boldsymbol{\gamma})} + \boldsymbol{\rho}^\top H^\top (MV^*+C)^\top \Phi^\top Z I^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}],\\
K^{(\boldsymbol{\beta},\lambda)} &= (K^{(\lambda,\boldsymbol{\beta})})^\top = -\Omega[I^{(\boldsymbol{\beta},\boldsymbol{\beta})} X^\top \Phi^2 V^* T\boldsymbol{\rho} + I^{(\boldsymbol{\beta},\boldsymbol{\gamma})} Z^\top \Phi(MV^*+C)H\boldsymbol{\rho}],\\
K^{(\boldsymbol{\gamma},\lambda)} &= (K^{(\lambda,\boldsymbol{\gamma})})^\top = -\Omega[I^{(\boldsymbol{\gamma},\boldsymbol{\beta})} X^\top \Phi^2 V^* T\boldsymbol{\rho} + I^{(\boldsymbol{\gamma},\boldsymbol{\gamma})} Z^\top \Phi(MV^*+C)H\boldsymbol{\rho}],\\
K^{(\lambda,\lambda)} &= \Omega,
\end{aligned}
$$

with

$$
\begin{aligned}
\Omega = \{&\boldsymbol{\rho}^\top \Phi^2 V^* \boldsymbol{\rho} - [\boldsymbol{\rho}^\top T^\top (V^*)^\top (\Phi^2)^\top X I^{(\boldsymbol{\beta},\boldsymbol{\beta})} X^\top \Phi^2 V^* T\boldsymbol{\rho} + \boldsymbol{\rho}^\top H^\top (MV^*\\
&+C)^\top \Phi^\top Z I^{(\boldsymbol{\gamma},\boldsymbol{\beta})} X^\top \Phi^2 V^* T\boldsymbol{\rho} + \boldsymbol{\rho}^\top T^\top (V^*)^\top (\Phi^2)^\top X I^{(\boldsymbol{\beta},\boldsymbol{\gamma})} Z^\top \Phi(MV^*+C)H\boldsymbol{\rho}\\
&+\boldsymbol{\rho}^\top H^\top (MV^*+C)^\top \Phi^\top Z I^{(\boldsymbol{\gamma},\boldsymbol{\gamma})} Z^\top \Phi(MV^*+C)H\boldsymbol{\rho}]\}^{-1}.
\end{aligned}
$$

2.2 NONLINEAR VARYING PRECISION BETA MODEL WITH PARAMETRIC MEAN LINK FUNCTION

Let $y_1,\ldots,y_n$ be a set of n independent random variables each following the beta law given in (2.1). The nonlinear beta regression model with parametric mean link function is given by

$$
\begin{aligned}
g(\mu_t,\lambda) &= f_1(\boldsymbol{x}_t^\top,\boldsymbol{\beta}) = \eta_{1t},\\
h(\phi_t) &= f_2(\boldsymbol{z}_t^\top,\boldsymbol{\gamma}) = \eta_{2t},
\end{aligned}
$$

where $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)^\top \in \mathbb{R}^p$ and $\boldsymbol{\gamma} = (\gamma_1,\ldots,\gamma_q)^\top \in \mathbb{R}^q$ are unknown parameters vectors of dimensions $p\times 1$ and $q\times 1$, respectively. The mean and precision nonlinear predictor vectors are, respectively, $\boldsymbol{\eta_1} = (\eta_{11},\ldots,\eta_{1n})^\top$ and $\boldsymbol{\eta_2} = (\eta_{21},\ldots,\eta_{2n})^\top$, $\boldsymbol{x}_t^\top = (x_{t1},\ldots,x_{tp_1})$ and $\boldsymbol{z}_t^\top = (z_{t1},\ldots,z_{tq_1})$ being the covariates vectors, $t = 1,\ldots,n$, $p_1 \le p$ and $q_1 \le q$, with $p+q+1 < n$. Here, $f_1(\cdot,\cdot)$ and $f_2(\cdot,\cdot)$ are differentiable and continuous functions, $g(\cdot,\cdot)$ and $h(\cdot)$ being defined as in the linear case. Also, we assume $\dfrac{\partial\eta_1}{\partial\boldsymbol{\beta}} = \mathcal{X}$ and $\dfrac{\partial\eta_2}{\partial\boldsymbol{\gamma}} = \mathcal{Z}$, in which $\mathcal{X}$ and $\mathcal{Z}$ have full column, rank p and q, respectively.

Taking the first order derivatives of the log-likelihood function in (2.11) with respect to each component of $\boldsymbol{\theta}$, we obtain the score vector, which is given in matrix form by

$$
\begin{aligned}
\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \mathcal{X}^\top \Phi T(\boldsymbol{y}^* - \boldsymbol{\mu}^*),\\
\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}) &= \mathcal{Z}^\top H\left[M(\boldsymbol{y}^* - \boldsymbol{\mu}^*) + (\boldsymbol{y}^\dagger - \boldsymbol{\mu}^\dagger)\right],
\end{aligned}
$$

$$\boldsymbol{U}_{\lambda}(\boldsymbol{\theta}) = \boldsymbol{\rho}^{\top}\Phi(\boldsymbol{y}^{*} - \boldsymbol{\mu}^{*}).$$

The elements of the Hessian matrix are

$$\begin{aligned}
\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\beta_i\partial\beta_r} &= \sum_{t=1}^{n}\left\{\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\beta_i\partial\beta_r}\right\},\\
&= \sum_{t=1}^{n}\left\{\frac{\partial}{\partial\beta_r}\left[\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\beta_i}\right]\right\},\\
&= \sum_{t=1}^{n}\left\{\frac{\partial}{\partial\beta_r}\left[\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial\mu_t}{\partial\eta_{1t}}\right]\frac{\partial\eta_{1t}}{\partial\beta_i} + \frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial}{\partial\beta_r}\left[\frac{\partial\eta_{1t}}{\partial\beta_i}\right]\right\},\\
&= \sum_{t=1}^{n}\left\{\frac{\partial}{\partial\mu_t}\left[\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial\mu_t}{\partial\eta_{1t}}\right]\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\beta_r}\frac{\partial\eta_{1t}}{\partial\beta_i} + \frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial^2\eta_{1t}}{\partial\beta_i\partial\beta_r}\right\},\\
&= \sum_{t=1}^{n}\left\{\left[\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\mu_t^2}\frac{\partial\mu_t}{\partial\eta_{1t}} + \frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial^2\mu_t}{\partial\eta_{1t}\partial\mu_t}\right]\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\beta_r}\frac{\partial\eta_{1t}}{\partial\beta_i} + \frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial^2\eta_{1t}}{\partial\beta_i\partial\beta_r}\right\},\\
\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\beta_i\partial\gamma_j} &= \sum_{t=1}^{n}\left\{\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\beta_i\partial\gamma_j}\right\},\\
&= \sum_{t=1}^{n}\left\{\frac{\partial}{\partial\gamma_j}\left[\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\beta_i}\right]\right\},\\
&= \sum_{t=1}^{n}\left\{\left[\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\mu_t\partial\gamma_j}\right]\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\beta_i}\right\},\\
&= \sum_{t=1}^{n}\left\{\left[\frac{\partial}{\partial\phi_t}\left(\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\right)\frac{\partial\phi_t}{\partial\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\gamma_j}\right]\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\beta_i}\right\},\\
&= \sum_{t=1}^{n}\left\{\left[\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\mu_t\partial\phi_t}\frac{\partial\phi_t}{\partial\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\gamma_j}\right]\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\beta_i}\right\},\\
\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\beta_i\partial\lambda} &= \sum_{t=1}^{n}\left\{\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\beta_i\partial\lambda}\right\},\\
&= \sum_{t=1}^{n}\left\{\frac{\partial}{\partial\lambda}\left[\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial\mu_t}{\partial\eta_{1t}}\frac{\partial\eta_{1t}}{\partial\beta_i}\right]\right\},\\
&= \sum_{t=1}^{n}\left\{\left[\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\mu_t^2}\frac{\partial\mu_t}{\partial\lambda}\frac{\partial\mu_t}{\partial\eta_{1t}} + \frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial}{\partial\lambda}\left(\frac{\partial\mu_t}{\partial\eta_{1t}}\right)\right]\frac{\partial\eta_{1t}}{\partial\beta_i}\right\},\\
\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\gamma_j\partial\gamma_g} &= \sum_{t=1}^{n}\left\{\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\gamma_j\partial\gamma_g}\right\},\\
&= \sum_{t=1}^{n}\left\{\frac{\partial}{\partial\gamma_g}\left[\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\phi_t}\frac{\partial\phi_t}{\partial\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\gamma_j}\right]\right\},\\
&= \sum_{t=1}^{n}\left\{\frac{\partial}{\partial\gamma_g}\left[\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\phi_t}\frac{\partial\phi_t}{\partial\eta_{2t}}\right]\frac{\partial\eta_{2t}}{\partial\gamma_j} + \frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\phi_t}\frac{\partial\phi_t}{\partial\eta_{2t}}\frac{\partial}{\partial\gamma_g}\left[\frac{\partial\eta_{2t}}{\partial\gamma_j}\right]\right\},\\
&= \sum_{t=1}^{n}\left\{\frac{\partial}{\partial\phi_t}\left[\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\phi_t}\frac{\partial\phi_t}{\partial\eta_{2t}}\right]\frac{\partial\phi_t}{\partial\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\gamma_g}\frac{\partial\eta_{2t}}{\partial\gamma_j} + \frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\phi_t}\frac{\partial\phi_t}{\partial\eta_{2t}}\frac{\partial^2\eta_{2t}}{\partial\gamma_g\partial\gamma_j}\right\},
\end{aligned}$$

$$= \sum_{t=1}^{n}\left\{\left[\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\phi_t^2}\frac{\partial\phi_t}{\partial\eta_{2t}}+\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\phi_t}\frac{\partial^2\phi_t}{\partial\eta_{2t}\partial\phi_t}\right]\frac{\partial\phi_t}{\partial\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\gamma_g}\frac{\partial\eta_{2t}}{\partial\gamma_j}+\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\phi_t}\frac{\partial\phi_t}{\partial\eta_{2t}}\frac{\partial^2\eta_{2t}}{\partial\gamma_g\partial\gamma_j}\right\},$$

$$\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\gamma_j\partial\lambda} = \sum_{t=1}^{n}\left\{\frac{\partial^2\ell(\mu_t,\phi_t)}{\partial\gamma_j\partial\lambda}\right\},$$

$$= \sum_{t=1}^{n}\left\{\frac{\partial}{\partial\lambda}\left[\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\phi_t}\frac{\partial\phi_t}{\partial\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\gamma_j}\right]\right\},$$

$$= \sum_{t=1}^{n}\left\{\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\phi_t\partial\mu_t}\frac{\partial\mu_t}{\partial\lambda}\frac{\partial\phi_t}{\partial\eta_{2t}}\frac{\partial\eta_{2t}}{\partial\gamma_j}\right\},$$

$$\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\lambda^2} = \sum_{t=1}^{n}\left\{\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\lambda^2}\right\},$$

$$= \sum_{t=1}^{n}\left\{\frac{\partial}{\partial\lambda}\left(\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial\mu_t}{\partial\lambda}\right)\right\},$$

$$= \sum_{t=1}^{n}\left\{\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\mu_t^2}\frac{\partial\mu_t}{\partial\lambda}\frac{\partial\mu_t}{\partial\lambda}+\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\frac{\partial^2\mu}{\partial\lambda^2}\right\}.$$

Therefore,

$$\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\beta_i\partial\beta_r} = \sum_{t=1}^{n}\left\{\left[-\phi_t^2\upsilon_t^*\left(\frac{1}{g'(\mu_t,\lambda)}\right)+\phi_t(y_t^*-\mu_t^*)\left(-\frac{g''(\mu_t,\lambda)}{g'(\mu_t,\lambda)^2}\right)\right]\left(\frac{1}{g'(\mu_t,\lambda)}\right)x_{tr}x_{ti}\right.$$
$$\left.+\phi_t(y_t^*-\mu_t^*)\left(\frac{1}{g'(\mu_t,\lambda)}\right)\frac{\partial^2\eta_{1t}}{\partial\beta_i\partial\beta_r}\right\},$$

$$\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\beta_i\partial\gamma_j} = \sum_{t=1}^{n}\left\{\left[(y_t^*-\mu_t^*)-\phi_t(\mu_t\upsilon_t^*+c_t^{*\dagger})\right]\left(\frac{1}{h'(\phi_t)}\right)z_{tj}\left(\frac{1}{g'(\mu_t,\lambda)}\right)x_{ti}\right\},$$

$$\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\beta_i\partial\lambda} = \sum_{t=1}^{n}\left\{\left[-\phi_t^2\upsilon_t^*\rho_t\left(\frac{1}{g'(\mu_t,\lambda)}\right)+\phi_t(y_t^*-\mu_t^*)\frac{\partial}{\partial\lambda}\left(\frac{1}{g'(\mu_t,\lambda)}\right)\right]x_{ti}\right\},$$

$$\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\gamma_j\partial\gamma_g} = \sum_{t=1}^{n}\left\{\left[\left(-\mu_t^2v_t^*-2\mu_tc_t^{*\dagger}-v^{\dagger}\right)\left(\frac{1}{h'(\phi_t)}\right)+\left(\mu_t(y_t^*-\mu_t^*)+(y_t^{\dagger}-\mu_t^{\dagger})\right)\left(-\frac{h''(\phi_t)}{h'(\phi_t)^2}\right)\right]\right.$$
$$\left.\times\left(\frac{1}{h'(\phi_t)}\right)z_{tg}z_{tj}\right\}+\sum_{t=1}^{n}\left\{\left(\mu_t(y_t^*-\mu_t^*)+(y_t^{\dagger}-\mu_t^{\dagger})\right)\frac{1}{h'(\phi_t)}\frac{\partial^2\eta_{2t}}{\partial\gamma_g\partial\gamma_j}\right\},$$

$$\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\gamma_j\partial\lambda} = \sum_{t=1}^{n}\left\{\left[(y_t^*-\mu_t^*)-\phi_t(\mu_tv_t^*+c_t^{*\dagger})\right]\rho_t\left(\frac{1}{h'(\phi_t)}\right)z_{tj}\right\},$$

$$\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\lambda^2} = \sum_{t=1}^{n}\left\{-\phi_t^2\upsilon_t^*\rho_t\rho_t+\phi_t(y_t^*-\mu_t^*)\frac{\partial^2\mu}{\partial\lambda^2}\right\}.$$

The observed information matrix then is given by

$$J \equiv J(\boldsymbol{\theta}) = \begin{pmatrix} J_{(\boldsymbol{\beta},\boldsymbol{\beta})} & J_{(\boldsymbol{\beta},\boldsymbol{\gamma})} & J_{(\boldsymbol{\beta},\lambda)} \\ J_{(\boldsymbol{\gamma},\boldsymbol{\beta})} & J_{(\boldsymbol{\gamma},\boldsymbol{\gamma})} & J_{(\boldsymbol{\gamma},\lambda)} \\ J_{(\lambda,\boldsymbol{\beta})} & J_{(\lambda,\boldsymbol{\gamma})} & J_{(\lambda,\lambda)} \end{pmatrix},$$

where $J_{(\boldsymbol{\beta},\boldsymbol{\beta})} = \mathcal{X}^\top[\Phi T V^* + S T^2(Y^* - M^*)]T\Phi\mathcal{X} + [(y^* - \mu^*)^\top \Phi T][\tilde{\mathcal{X}}]$, $J_{(\boldsymbol{\beta},\boldsymbol{\gamma})} = J^\top_{(\boldsymbol{\gamma},\boldsymbol{\beta})} = -\mathcal{X}^\top[(Y^* - M^*) - \Phi(MV^* + C)]TH\mathcal{Z}$, $J_{(\boldsymbol{\beta},\lambda)} = J^\top_{(\lambda,\boldsymbol{\beta})} = \mathcal{X}^\top[\Phi^2 V^* T\boldsymbol{\rho} - \Phi(Y^* - M^*)\boldsymbol{w}]$, $J_{(\boldsymbol{\gamma},\boldsymbol{\gamma})} = \mathcal{Z}^\top\{H(M^2V^* + 2MC + V^\dagger) + [M(Y^* - M^*) + (Y^\dagger - M^\dagger)]H^2 Q\}H\mathcal{Z} + [(y^* - \mu^*)^\top M + (y^\dagger - \mu^\dagger)^\top H][\tilde{\mathcal{Z}}]$, $J_{(\boldsymbol{\gamma},\lambda)} = J^\top_{(\lambda,\boldsymbol{\gamma})} = -\mathcal{Z}^\top[(Y^* - M^*) - \Phi(MV^* + C)]H\boldsymbol{\rho}$, and $J_{(\lambda,\lambda)} = [\Phi^2 V^* \boldsymbol{\rho}^2 - \Phi(Y^* - M^*)\boldsymbol{\varrho}]^\top \boldsymbol{\imath}$. Here, $[\cdot][\cdot]$ represents the bracket product between a matrix and an array as defined in Wei (1998, p. 188), and $\dfrac{\partial^2 \eta_{1t}}{\partial \beta_i \partial \beta_r} = \tilde{\mathcal{X}}$ and $\dfrac{\partial^2 \eta_{2t}}{\partial \gamma_j \partial \gamma_u} = \tilde{\mathcal{Z}}$ are arrays of dimensions $n \times p \times p$ and $n \times q \times q$, respectively.

Since $\mathbb{E}\left(\dfrac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t}\right) = \mathbb{E}\left(\dfrac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t}\right) = 0$, Fisher's information matrix is given by

$$K \equiv K(\boldsymbol{\theta}) = \begin{pmatrix} K_{(\boldsymbol{\beta},\boldsymbol{\beta})} & K_{(\boldsymbol{\beta},\boldsymbol{\gamma})} & K_{(\boldsymbol{\beta},\lambda)} \\ K_{(\boldsymbol{\gamma},\boldsymbol{\beta})} & K_{(\boldsymbol{\gamma},\boldsymbol{\gamma})} & K_{(\boldsymbol{\gamma},\lambda)} \\ K_{(\lambda,\boldsymbol{\beta})} & K_{(\lambda,\boldsymbol{\gamma})} & K_{(\lambda,\lambda)} \end{pmatrix},$$

where $K_{(\boldsymbol{\beta},\boldsymbol{\beta})} = \mathcal{X}^\top \Phi^2 V^* T^2 \mathcal{X}$, $K_{(\boldsymbol{\beta},\boldsymbol{\gamma})} = K^\top_{(\boldsymbol{\gamma},\boldsymbol{\beta})} = \mathcal{X}^\top \Phi(MV^* + C)TH\mathcal{Z}$, $K_{(\boldsymbol{\beta},\lambda)} = K^\top_{(\lambda,\boldsymbol{\beta})} = \mathcal{X}^\top \Phi^2 V^* T \boldsymbol{\rho}$, $K_{(\boldsymbol{\gamma},\boldsymbol{\gamma})} = \mathcal{Z}^\top H(M^2 V^* + 2MC + V^\dagger)H\mathcal{Z}$, $K_{(\boldsymbol{\gamma},\lambda)} = K^\top_{(\lambda,\boldsymbol{\gamma})} = \mathcal{Z}^\top \Phi(MV^* + C)H\boldsymbol{\rho}$, and $K_{(\lambda,\lambda)} = \boldsymbol{\rho}^\top \Phi^2 V^* \boldsymbol{\rho}$.

It can be shown that the formulas for the score vector and Fisher's information matrix for the nonlinear case are equal to those obtained in the linear case after we replace X and Z with $\mathcal{X}$ and $\mathcal{Z}$, respectively. Therefore, Fisher's information inverse matrix for the nonlinear varying precision beta regression model with parametric mean link function is equal to that in (2.23), with X and Z replaced by $\mathcal{X}$ and $\mathcal{Z}$, respectively.

3 IMPROVED LIKELIHOOD TESTING INFERENCES

In this chapter, we shall develop improved likelihood testing inference for the linear and nonlinear varying precision beta regression models with parametric mean link function. Our goal is to obtain testing criteria that can be used to achieve good control of the type I error probability when the number of observations is small.

3.1 THE LIKELIHOOD RATIO TEST

Let $\boldsymbol{\theta} = (\boldsymbol{\nu}^\top, \boldsymbol{\delta}^\top)^\top$, where $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_l)^\top$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_s)^\top$ are the parameter of interest and the nuisance parameter, respectively. Notice that $l + s = p + q + 1$. Our interest lies in testing $\mathcal{H}_0 : \boldsymbol{\nu} = \boldsymbol{\nu}_0$ against $\mathcal{H}_1 : \boldsymbol{\nu} \neq \boldsymbol{\nu}_0$, where $\boldsymbol{\nu}_0$ is a given l-vector. Let $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\nu}}^\top, \hat{\boldsymbol{\delta}}^\top)^\top$ and $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\nu}_0^\top, \tilde{\boldsymbol{\delta}}^\top)^\top$ be the unrestricted and restricted maximum likelihood estimators of $\boldsymbol{\theta}$, respectively. The likelihood test statistic is

$$w = 2[\ell(\hat{\boldsymbol{\nu}}, \hat{\boldsymbol{\delta}}) - \ell(\boldsymbol{\nu}_0, \tilde{\boldsymbol{\delta}})].$$

In large samples and under $\mathcal{H}_0$, w is approximated distributed as χ^2_l (PAWITAN, 2001, Chapter 9). The test is then carried out using critical values obtained from such a limiting null distribution and, as result, size distortions are likely to occur when the sample size is small.

3.2 SKOVGAARD'S ADJUSTMENTS FOR THE LIKELIHOOD RATIO TEST STATISTIC

A variant of the likelihood test statistic designed to deliver more accurate inferences in finite samples was developed by Skovgaard (2001) building upon previous results obtained by Barndorff-Nielsen (1986), Barndorff-Nielsen (1991). It is given by

$$w^* = w - 2\log(\xi).$$

An alternative modified test statistic is given by

$$w^{**} = w\left[1 - w^{-1}\log(\xi)\right]^2.$$

An advantage of the latter is that it is always non-negative. Here,

$$\xi = \frac{(|\tilde{K}||\hat{K}||\tilde{J}_{\boldsymbol{\delta},\boldsymbol{\delta}}|)^{1/2}}{|\bar{\Upsilon}||(\tilde{K}\bar{\Upsilon}^{-1}\hat{J}\hat{K}^{-1}\bar{\Upsilon})_{\boldsymbol{\delta},\boldsymbol{\delta}}|^{1/2}} \frac{(\tilde{\boldsymbol{U}}^\top\bar{\Upsilon}^{-1}\hat{K}\hat{J}^{-1}\bar{\Upsilon}\tilde{K}^{-1}\tilde{\boldsymbol{U}})^{l/2}}{w^{l/2-1}\tilde{\boldsymbol{U}}^\top\bar{\Upsilon}^{-1}\bar{\boldsymbol{q}}},$$

where hat and tilde denote, respectively, evaluation at the unrestricted and restricted maximum likelihood estimators. Here, J is the observed information matrix, K is Fisher's information, $\boldsymbol{U}$

is the score vector, and $J_{\boldsymbol{\delta},\boldsymbol{\delta}}$ denotes the $s \times s$ observed information matrix corresponding to $\boldsymbol{\delta}$. Likewise, the subscript $\boldsymbol{\delta},\boldsymbol{\delta}$ when applied to a matrix indicates its block relative to $\boldsymbol{\delta}$. Additionally, $\bar{\boldsymbol{q}}$ is an $l+s$ column vector and $\bar{\Upsilon}$ is an $(l+s)\times(l+s)$ matrix which come, respectively, from

$$\boldsymbol{q} = \mathbb{E}_{\boldsymbol{\theta}_1}\left[\boldsymbol{U}(\boldsymbol{\theta_1})(\ell(\boldsymbol{\theta_1}) - \ell(\boldsymbol{\theta}))\right] \quad \text{and} \quad \Upsilon = \mathbb{E}_{\boldsymbol{\theta}_1}\left[\boldsymbol{U}(\boldsymbol{\theta_1})\boldsymbol{U}^\top(\boldsymbol{\theta})\right]$$

by replacing $\boldsymbol{\theta_1}$ with $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$ with $\tilde{\boldsymbol{\theta}}$ after expected values are computed.

The two modified test statistics are invariant under reparametrizations of the form $(\nu,\delta)\mapsto(\nu,\varphi(\nu,\delta))$ and the null distributions of w^* and w^{**} converge to χ^2_l as $n\to\infty$. For further details, see Skovgaard (2001).

Our interest lies in obtaining the quantities that define the two modified likelihood ratio test statistics in the context of the varying precision beta regression model with parametric mean link function and write them in matrix form that can be easily numerically computed by practitioners. In what follows, we provide details on the derivation of the corrected likelihood test statistics for that class of models.

In the linear case, observe that $\bar{\boldsymbol{q}}$ is obtained from

$$\boldsymbol{q} = \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}(\boldsymbol{\theta_1})(\ell(\boldsymbol{\theta_1}) - \ell(\boldsymbol{\theta}))] = \begin{bmatrix} \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta_1})\ell(\boldsymbol{\theta_1})] - \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta_1})\ell(\boldsymbol{\theta})] \\ \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta_1})\ell(\boldsymbol{\theta_1})] - \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta_1})\ell(\boldsymbol{\theta})] \\ \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\lambda}(\boldsymbol{\theta_1})\ell(\boldsymbol{\theta_1})] - \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\lambda}(\boldsymbol{\theta_1})\ell(\boldsymbol{\theta})] \end{bmatrix}.$$

From (2.11) and (2.18) we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta})\ell(\boldsymbol{\theta})] &= \mathbb{E}_{\boldsymbol{\theta}}\left\{X^\top\Phi T(\boldsymbol{y}^* - \boldsymbol{\mu}^*)[(\boldsymbol{y}^* - \boldsymbol{\mu}^*)^\top(\Phi M - I) + (\boldsymbol{y}^\dagger - \boldsymbol{\mu}^\dagger)^\top(\Phi - 2I) + \boldsymbol{b}^\top]\boldsymbol{\imath}\right\} \\
&= X^\top\Phi T\left\{\mathbb{E}_{\boldsymbol{\theta}}[(\boldsymbol{y}^* - \boldsymbol{\mu}^*)(\boldsymbol{y}^* - \boldsymbol{\mu}^*)^\top](\Phi M - I) + \mathbb{E}_{\boldsymbol{\theta}}[(\boldsymbol{y}^* - \boldsymbol{\mu}^*)(\boldsymbol{y}^\dagger - \boldsymbol{\mu}^\dagger)^\top]\right. \\
&\quad \left.\times(\Phi - 2I) + \mathbb{E}_{\boldsymbol{\theta}}[(\boldsymbol{y}^* - \boldsymbol{\mu}^*)]\boldsymbol{b}^\top\right\}\boldsymbol{\imath} \\
&= X^\top\Phi T\left\{V^*(\Phi M - I) + C(\Phi - 2I)\right\}\boldsymbol{\imath}.
\end{aligned}$$

For all $t\neq u$, y_t and y_u are independent, and $\mathbb{E}_{\theta_1}(y_t^* - \mu_t^{*(1)}) = 0$. Therefore, $\mathbb{E}_{\theta_1}[(y_t^* - \mu_t^{*(1)})(y_u^* - \mu_u^*)] = 0$ and $\mathbb{E}_{\theta_1}[(y_t^* - \mu_t^{*(1)})(y_t^* - \mu_t^*)] = \mathbb{E}_{\theta_1}[(y_t^* - \mu_t^{*(1)})(y_t^* - \mu_t^{*(1)})] + \mathbb{E}_{\theta_1}[(y_t^* - \mu_t^{*(1)})(\mu_t^{*(1)} - \mu_t^*)] = \mathbb{E}_{\theta_1}[(y_t^* - \mu_t^{*(1)})^2] = \upsilon_t^{*(1)}$. Evaluation at θ_1 is indicated by the superscript '(1)'.

After some algebra, we obtain

$$\mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta_1})\ell(\boldsymbol{\theta})] = X^\top\Phi^{(1)}T^{(1)}\left\{V^{*(1)}(\Phi M - I) + C^{(1)}(\Phi - 2I)\right\}\boldsymbol{\imath}.$$

Thus,

$$\mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta_1})\ell(\boldsymbol{\theta_1})] - \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta_1})\ell(\boldsymbol{\theta})] = X^\top\Phi^{(1)}T^{(1)}\left\{V^{*(1)}(\Phi^{(1)}M^{(1)} - I) + C^{(1)}(\Phi^{(1)} - 2I)\right\}\boldsymbol{\imath}$$

$$
\begin{aligned}
&- X^\top \Phi^{(1)} T^{(1)} \big\{ V^{*(1)}(\Phi M - I) + C^{(1)}(\Phi - 2I) \big\} \boldsymbol{\iota} \\
&= X^\top \Phi^{(1)} T^{(1)} \big\{ V^{*(1)}(\Phi^{(1)} M^{(1)} - \Phi M) + C^{(1)}(\Phi^{(1)} \\
&- \Phi) \big\} \boldsymbol{\iota}.
\end{aligned}
$$

Using Equations (2.11) and (2.19), we obtain

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta})\ell(\boldsymbol{\theta})] &= \mathbb{E}_{\boldsymbol{\theta}} \left\{ Z^\top H[M(\boldsymbol{y}^* - \boldsymbol{\mu}^*) + (\boldsymbol{y}^\dagger - \boldsymbol{\mu}^\dagger)][(\boldsymbol{y}^* - \boldsymbol{\mu}^*)^\top (\Phi M - I) + (\boldsymbol{y}^\dagger - \boldsymbol{\mu}^\dagger)^\top (\Phi \right. \\
&\left. - 2I) + \boldsymbol{b}^\top] \boldsymbol{\iota} \right\} \\
&= Z^\top H \big\{ M \mathbb{E}_{\boldsymbol{\theta}}[(\boldsymbol{y}^* - \boldsymbol{\mu}^*)(\boldsymbol{y}^* - \boldsymbol{\mu}^*)^\top](\Phi M - I) + M \mathbb{E}_{\boldsymbol{\theta}}[(\boldsymbol{y}^* - \boldsymbol{\mu}^*)(\boldsymbol{y}^\dagger - \boldsymbol{\mu}^\dagger)^\top](\Phi \\
&- 2I) + \mathbb{E}_{\boldsymbol{\theta}}[(\boldsymbol{y}^\dagger - \boldsymbol{\mu}^\dagger)(\boldsymbol{y}^* - \boldsymbol{\mu}^*)^\top](\Phi M - I) + \mathbb{E}_{\boldsymbol{\theta}}[(\boldsymbol{y}^\dagger - \boldsymbol{\mu}^\dagger)(\boldsymbol{y}^\dagger - \boldsymbol{\mu}^\dagger)^\top](\Phi - 2I) \big\} \boldsymbol{\iota} \\
&= Z^\top H \left\{ M V^* (\Phi M - I) + M C (\Phi - 2I) + C(\Phi M - I) + V^\dagger (\Phi - 2I) \right\} \boldsymbol{\iota} \\
&= Z^\top H \left\{ (M V^* + C)(\Phi M - I) + (M C + V^\dagger)(\Phi - 2I) \right\} \boldsymbol{\iota}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}_1)\ell(\boldsymbol{\theta}_1)] - \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}_1)\ell(\boldsymbol{\theta})] &= Z^\top H^{(1)} \big\{ (M^{(1)} V^{*(1)} + C^{(1)})(\Phi^{(1)} M^{(1)} - I) \\
&+ (M^{(1)} C^{(1)} + V^{\dagger(1)})(\Phi^{(1)} - 2I) \big\} \boldsymbol{\iota} \\
&- Z^\top H^{(1)} \big\{ (M^{(1)} V^{*(1)} + C^{(1)})(\Phi M - I) + (M^{(1)} C^{(1)} \\
&+ V^{\dagger(1)})(\Phi - 2I) \big\} \boldsymbol{\iota} \\
&= Z^\top H^{(1)} \big\{ (M^{(1)} V^{*(1)} + C^{(1)})(\Phi^{(1)} M^{(1)} - \Phi M) \\
&+ (M^{(1)} C^{(1)} + V^{\dagger(1)}) \times (\Phi^{(1)} - \Phi) \big\} \boldsymbol{\iota}.
\end{aligned}
$$

Similarly, from (2.11) and (2.20) it follows that

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{U}_\lambda(\boldsymbol{\theta})\ell(\boldsymbol{\theta})] &= \mathbb{E}_{\boldsymbol{\theta}} \big\{ \boldsymbol{\rho}^\top \Phi (\boldsymbol{y}^* - \boldsymbol{\mu}^*)[(\boldsymbol{y}^* - \boldsymbol{\mu}^*)^\top (\Phi M - I) + (\boldsymbol{y}^\dagger - \boldsymbol{\mu}^\dagger)^\top (\Phi - 2I) + \boldsymbol{b}^\top] \boldsymbol{\iota} \big\} \\
&= \boldsymbol{\rho}^\top \Phi \big\{ \mathbb{E}_{\boldsymbol{\theta}}[(\boldsymbol{y}^* - \boldsymbol{\mu}^*)(\boldsymbol{y}^* - \boldsymbol{\mu}^*)^\top](\Phi M - I) + \mathbb{E}_{\boldsymbol{\theta}}[(\boldsymbol{y}^* - \boldsymbol{\mu}^*)(\boldsymbol{y}^\dagger - \boldsymbol{\mu}^\dagger)^\top](\Phi - 2I) \big\} \boldsymbol{\iota} \\
&= \boldsymbol{\rho}^\top \Phi \big\{ V^*(\Phi M - I) + C(\Phi - 2I) \big\} \boldsymbol{\iota}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_\lambda(\boldsymbol{\theta}_1)\ell(\boldsymbol{\theta}_1)] - \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_\lambda(\boldsymbol{\theta}_1)\ell(\boldsymbol{\theta})] &= \boldsymbol{\rho}^{\top(1)} \Phi^{(1)} \big\{ V^{*(1)}(\Phi^{(1)} M^{(1)} - I) + C^{(1)}(\Phi^{(1)} - 2I) \big\} \boldsymbol{\iota} \\
&- \boldsymbol{\rho}^{\top(1)} \Phi^{(1)} \big\{ V^{*(1)}(\Phi M - I) + C^{(1)}(\Phi - 2I) \big\} \boldsymbol{\iota} \\
&= \boldsymbol{\rho}^{\top(1)} \Phi^{(1)} \big\{ V^{*(1)}(\Phi^{(1)} M^{(1)} - \Phi M) + C^{(1)}(\Phi^{(1)} - \Phi) \big\} \boldsymbol{\iota}.
\end{aligned}
$$

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

$$\Upsilon = \begin{bmatrix} \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\boldsymbol{\beta}}^\top(\boldsymbol{\theta})] & \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\boldsymbol{\gamma}}^\top(\boldsymbol{\theta})] & \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\lambda}^\top(\boldsymbol{\theta})] \\ \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\boldsymbol{\beta}}^\top(\boldsymbol{\theta})] & \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\boldsymbol{\gamma}}^\top(\boldsymbol{\theta})] & \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\lambda}^\top(\boldsymbol{\theta})] \\ \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\lambda}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\boldsymbol{\beta}}^\top(\boldsymbol{\theta})] & \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\lambda}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\boldsymbol{\gamma}}^\top(\boldsymbol{\theta})] & \mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\lambda}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\lambda}^\top(\boldsymbol{\theta})] \end{bmatrix}.$$

From Equations (2.18), (2.19), and (2.20) we obtain

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\boldsymbol{\beta}}^\top(\boldsymbol{\theta})] &= \mathbb{E}_{\boldsymbol{\theta}_1}\{X^\top\Phi^{(1)}T^{(1)}(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})[X^\top\Phi T(\boldsymbol{y}^*-\boldsymbol{\mu}^*)]^\top\} \\
&= \mathbb{E}_{\boldsymbol{\theta}_1}[X^\top\Phi^{(1)}T^{(1)}(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^*-\boldsymbol{\mu}^*)^\top T\Phi X] \\
&= X^\top\Phi^{(1)}T^{(1)}\mathbb{E}_{\boldsymbol{\theta}_1}[(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^*-\boldsymbol{\mu}^*)^\top]T\Phi X \\
&= X^\top\Phi^{(1)}T^{(1)}V^{*(1)}T\Phi X, \\
\mathbb{E}_{\boldsymbol{\theta}_1}\big[\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\boldsymbol{\gamma}}^\top(\boldsymbol{\theta})\big] &= \mathbb{E}_{\boldsymbol{\theta}_1}\big\{X^\top\Phi^{(1)}T^{(1)}(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})\big[Z^\top H\big[M(\boldsymbol{y}^*-\boldsymbol{\mu}^*)+(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^\dagger)\big]\big]^\top\big\} \\
&= \mathbb{E}_{\boldsymbol{\theta}_1}\big[X^\top\Phi^{(1)}T^{(1)}(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})\big[(\boldsymbol{y}^*-\boldsymbol{\mu}^*)^\top M+(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^\dagger)^\top\big]HZ\big] \\
&= X^\top\Phi^{(1)}T^{(1)}\mathbb{E}_{\boldsymbol{\theta}_1}\big[(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})\big[(\boldsymbol{y}^*-\boldsymbol{\mu}^*)^\top M+(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^\dagger)^\top\big]\big]HZ \\
&= X^\top\Phi^{(1)}T^{(1)}\mathbb{E}_{\boldsymbol{\theta}_1}\big[(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^*-\boldsymbol{\mu}^*)^\top M+(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^\dagger)^\top\big]HZ \\
&= X^\top\Phi^{(1)}T^{(1)}\big\{V^{*(1)}M+C^{(1)}\big\}HZ, \\
\mathbb{E}_{\boldsymbol{\theta}_1}\big[\boldsymbol{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\lambda}^\top(\boldsymbol{\theta})\big] &= \mathbb{E}_{\boldsymbol{\theta}_1}\big\{X^\top\Phi^{(1)}T^{(1)}(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})\big[\boldsymbol{\rho}^\top\Phi(\boldsymbol{y}^*-\boldsymbol{\mu}^*)\big]^\top\big\} \\
&= \mathbb{E}_{\boldsymbol{\theta}_1}\big[X^\top\Phi^{(1)}T^{(1)}(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^*-\boldsymbol{\mu}^*)^\top\Phi\boldsymbol{\rho}\big] \\
&= X^\top\Phi^{(1)}T^{(1)}\mathbb{E}_{\boldsymbol{\theta}_1}\big[(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^*-\boldsymbol{\mu}^*)^\top\big]\Phi\boldsymbol{\rho} \\
&= X^\top\Phi^{(1)}T^{(1)}V^{*(1)}\Phi\boldsymbol{\rho}, \\
\mathbb{E}_{\boldsymbol{\theta}_1}\big[\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\boldsymbol{\beta}}^\top(\boldsymbol{\theta})\big] &= \mathbb{E}_{\boldsymbol{\theta}_1}\big\{Z^\top H^{(1)}\big[M^{(1)}(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})+(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^{\dagger(1)})\big]\big[X^\top\Phi T(\boldsymbol{y}^*-\boldsymbol{\mu}^*)\big]^\top\big\} \\
&= \mathbb{E}_{\boldsymbol{\theta}_1}\big[Z^\top H^{(1)}\big[M^{(1)}(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})+(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^{\dagger(1)})\big](\boldsymbol{y}^*-\boldsymbol{\mu}^*)^\top\, T\Phi X\big] \\
&= Z^\top H^{(1)}\mathbb{E}_{\boldsymbol{\theta}_1}\big[\big[M^{(1)}(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})+(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^{\dagger(1)})\big](\boldsymbol{y}^*-\boldsymbol{\mu}^*)^\top\big]T\Phi X \\
&= Z^\top H^{(1)}\mathbb{E}_{\boldsymbol{\theta}_1}\big[M^{(1)}(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^*-\boldsymbol{\mu}^*)^\top+(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^{\dagger(1)})(\boldsymbol{y}^*-\boldsymbol{\mu}^*)^\top\big]T\Phi X \\
&= Z^\top H^{(1)}\big\{M^{(1)}V^{*(1)}+C^{(1)}\big\}T\Phi X, \\
\mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\boldsymbol{\gamma}}^\top(\boldsymbol{\theta})] &= \mathbb{E}_{\boldsymbol{\theta}_1}\{Z^\top H^{(1)}[M^{(1)}(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})+(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^{\dagger(1)})][Z^\top H[M(\boldsymbol{y}^*-\boldsymbol{\mu}^*)+(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^\dagger)]]^\top\} \\
&= \mathbb{E}_{\boldsymbol{\theta}_1}[Z^\top H^{(1)}[M^{(1)}(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})+(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^{\dagger(1)})][(\boldsymbol{y}^*-\boldsymbol{\mu}^*)^\top M+(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^\dagger)^\top]HZ] \\
&= Z^\top H^{(1)}\mathbb{E}_{\boldsymbol{\theta}_1}[[M^{(1)}(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})+(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^{\dagger(1)})][(\boldsymbol{y}^*-\boldsymbol{\mu}^*)^\top M+(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^\dagger)^\top]]HZ \\
&= Z^\top H^{(1)}\mathbb{E}_{\boldsymbol{\theta}_1}[M^{(1)}(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^*-\boldsymbol{\mu}^*)^\top M+M^{(1)}(\boldsymbol{y}^*-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^\dagger)^\top \\
&\quad+(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^{\dagger(1)})(\boldsymbol{y}^*-\boldsymbol{\mu}^*)^\top M+(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^{\dagger(1)})(\boldsymbol{y}^\dagger-\boldsymbol{\mu}^\dagger)^\top]HZ \\
&= Z^\top H^{(1)}\big\{M^{(1)}V^{*(1)}M+(M^{(1)}+M)C^{(1)}+V^{\dagger(1)}\big\}HZ,
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}_1}\left[\boldsymbol{U}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\lambda}^{\top}(\boldsymbol{\theta})\right] &= \mathbb{E}_{\boldsymbol{\theta}_1}\left\{Z^{\top}H^{(1)}\left[M^{(1)}(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*(1)})+(\boldsymbol{y}^{\dagger}-\boldsymbol{\mu}^{\dagger(1)})\right]\left[\boldsymbol{\rho}^{\top}\Phi(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*})\right]^{\top}\right\}\\
&= \mathbb{E}_{\boldsymbol{\theta}_1}\left[Z^{\top}H^{(1)}\left[M^{(1)}(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*(1)})+(\boldsymbol{y}^{\dagger}-\boldsymbol{\mu}^{\dagger(1)})\right](\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*})^{\top}\Phi\boldsymbol{\rho}\right]\\
&= Z^{\top}H^{(1)}\mathbb{E}_{\boldsymbol{\theta}_1}\left[\left[M^{(1)}(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*(1)})+(\boldsymbol{y}^{\dagger}-\boldsymbol{\mu}^{\dagger(1)})\right](\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*})^{\top}\right]\Phi\boldsymbol{\rho}\\
&= Z^{\top}H^{(1)}\mathbb{E}_{\boldsymbol{\theta}_1}\left[M^{(1)}(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*})^{\top}+(\boldsymbol{y}^{\dagger}-\boldsymbol{\mu}^{\dagger(1)})(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*})^{\top}\right]\Phi\boldsymbol{\rho}\\
&= Z^{\top}H^{(1)}\left\{M^{(1)}V^{*(1)}+C^{(1)}\right\}\Phi\boldsymbol{\rho},\\
\mathbb{E}_{\boldsymbol{\theta}_1}\left[\boldsymbol{U}_{\lambda}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\boldsymbol{\beta}}^{\top}(\boldsymbol{\theta})\right] &= \mathbb{E}_{\boldsymbol{\theta}_1}\left\{\boldsymbol{\rho}^{\top(1)}\Phi^{(1)}(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*(1)})\left[X^{\top}\Phi T(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*})\right]^{\top}\right\}\\
&= \mathbb{E}_{\boldsymbol{\theta}_1}\left[\boldsymbol{\rho}^{\top(1)}\Phi^{(1)}(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*})^{\top}T\Phi X\right]\\
&= \boldsymbol{\rho}^{\top(1)}\Phi^{(1)}\mathbb{E}_{\boldsymbol{\theta}_1}\left[(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*})^{\top}\right]T\Phi X\\
&= \boldsymbol{\rho}^{\top(1)}\Phi^{(1)}V^{*(1)}T\Phi X,\\
\mathbb{E}_{\boldsymbol{\theta}_1}\left[\boldsymbol{U}_{\lambda}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\boldsymbol{\gamma}}^{\top}(\boldsymbol{\theta})\right] &= \mathbb{E}_{\boldsymbol{\theta}_1}\left\{\boldsymbol{\rho}^{\top(1)}\Phi^{(1)}(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*(1)})\left[Z^{\top}H\left[M(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*})+(\boldsymbol{y}^{\dagger}-\boldsymbol{\mu}^{\dagger})\right]\right]^{\top}\right\}\\
&= \mathbb{E}_{\boldsymbol{\theta}_1}\left[\boldsymbol{\rho}^{\top(1)}\Phi^{(1)}(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*(1)})\left[(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*})^{\top}M+(\boldsymbol{y}^{\dagger}-\boldsymbol{\mu}^{\dagger})^{\top}\right]HZ\right]\\
&= \boldsymbol{\rho}^{\top(1)}\Phi^{(1)}\mathbb{E}_{\boldsymbol{\theta}_1}\left[(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*(1)})\left[(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*})^{\top}M+(\boldsymbol{y}^{\dagger}-\boldsymbol{\mu}^{\dagger})^{\top}\right]\right]HZ\\
&= \boldsymbol{\rho}^{\top(1)}\Phi^{(1)}\mathbb{E}_{\boldsymbol{\theta}_1}\left[(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*})^{\top}M+(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^{\dagger}-\boldsymbol{\mu}^{\dagger})^{\top}\right]HZ\\
&= \boldsymbol{\rho}^{\top(1)}\Phi^{(1)}\left\{V^{*(1)}M+C^{(1)}\right\}HZ,\\
\mathbb{E}_{\boldsymbol{\theta}_1}[\boldsymbol{U}_{\lambda}(\boldsymbol{\theta}_1)\boldsymbol{U}_{\lambda}^{\top}(\boldsymbol{\theta})] &= \mathbb{E}_{\boldsymbol{\theta}_1}\left\{\boldsymbol{\rho}^{\top(1)}\Phi^{(1)}(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*(1)})[\boldsymbol{\rho}^{\top}\Phi(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*})]^{\top}\right\}\\
&= \mathbb{E}_{\boldsymbol{\theta}_1}\left[\boldsymbol{\rho}^{\top(1)}\Phi^{(1)}(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*})^{\top}\Phi\boldsymbol{\rho}\right]\\
&= \boldsymbol{\rho}^{\top(1)}\Phi^{(1)}\mathbb{E}_{\boldsymbol{\theta}_1}\left[(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*(1)})(\boldsymbol{y}^{*}-\boldsymbol{\mu}^{*})^{\top}\right]\Phi\boldsymbol{\rho}\\
&= \boldsymbol{\rho}^{\top(1)}\Phi^{(1)}V^{*(1)}\Phi\boldsymbol{\rho}.
\end{aligned}
$$

Finally, by combining the above results we can write $\bar{\boldsymbol{q}}$ and $\bar{\Upsilon}$ as

$$
\bar{\boldsymbol{q}} = \begin{bmatrix}
X^{\top}\Phi^{(1)}T^{(1)}\left[V^{*(1)}(\Phi^{(1)}M^{(1)}-\Phi M)+C^{(1)}(\Phi^{(1)}-\Phi)\right]\boldsymbol{\imath}\\
Z^{\top}H^{(1)}\left[(M^{(1)}V^{*(1)}+C^{(1)})(\Phi^{(1)}M^{(1)}-\Phi M)+(M^{(1)}C^{(1)}+V^{\dagger(1)})(\Phi^{(1)}-\Phi)\right]\boldsymbol{\imath}\\
\boldsymbol{\rho}^{\top(1)}\Phi^{(1)}\left[V^{*(1)}(\Phi^{(1)}M^{(1)}-\Phi M)+C^{(1)}(\Phi^{(1)}-\Phi)\right]\boldsymbol{\imath}
\end{bmatrix}
$$

and

$$
\bar{\Upsilon} = \begin{bmatrix} D_1 & D_2 & D_3\\ D_4 & D_5 & D_6\\ D_7 & D_8 & D_9 \end{bmatrix},
$$

where $D_1 = X^{\top}\Phi^{(1)}T^{(1)}V^{*(1)}T\Phi X$, $D_2 = X^{\top}\Phi^{(1)}T^{(1)}(V^{*(1)}M+C^{(1)}]HZ$, $D_3 = X^{\top}\Phi^{(1)}T^{(1)}V^{*(1)}\Phi\boldsymbol{\rho}$, $D_4 = Z^{\top}H^{(1)}(M^{(1)}V^{*(1)}+C^{(1)})T\Phi X$, $D_5 = Z^{\top}H^{(1)}[M^{(1)}V^{*(1)}M+(M^{(1)}+M)C^{(1)}+V^{\dagger(1)}]HZ$,

$D_6 = Z^\top H^{(1)}(M^{(1)}V^{*(1)} + C^{(1)})\Phi\boldsymbol{\rho}$, $D_7 = \boldsymbol{\rho}^{\top(1)}\Phi^{(1)}V^{*(1)}T\Phi X$, $D_8 = \boldsymbol{\rho}^{\top(1)}\Phi^{(1)}(V^{*(1)}M + C^{(1)})HZ$, and $D_9 = \boldsymbol{\rho}^{\top(1)}\Phi^{(1)}V^{*(1)}\Phi\boldsymbol{\rho}$.

The above expressions for $\bar{\boldsymbol{q}}$ and $\bar{\Upsilon}$ hold for models in which the precision is allowed to vary across observations. When precision is taken to be fixed, i.e., $\phi_1 = \cdots = \phi_n = \phi$, some simplifications take place, namely: the matrix Z becomes a column vector of ones, H reduces to the identity matrix and Φ equals the identity matrix multiplied by ϕ. We also note that under fixed precision Q becomes a matrix of zeros.

In the nonlinear case, $\bar{\boldsymbol{q}}$ and $\bar{\Upsilon}$ are given by

$$\bar{\boldsymbol{q}} = \begin{bmatrix} \mathcal{X}^\top\Phi^{(1)}T^{(1)}\left[V^{*(1)}(\Phi^{(1)}M^{(1)} - \Phi M) + C^{(1)}(\Phi^{(1)} - \Phi)\right]\boldsymbol{\iota} \\ \mathcal{Z}^\top H^{(1)}\left[(M^{(1)}V^{*(1)} + C^{(1)})(\Phi^{(1)}M^{(1)} - \Phi M) + (M^{(1)}C^{(1)} + V^{\dagger(1)})(\Phi^{(1)} - \Phi)\right]\boldsymbol{\iota} \\ \boldsymbol{\rho}^{\top(1)}\Phi^{(1)}\left[V^{*(1)}(\Phi^{(1)}M^{(1)} - \Phi M) + C^{(1)}(\Phi^{(1)} - \Phi)\right]\boldsymbol{\iota} \end{bmatrix}$$

and

$$\bar{\Upsilon} = \begin{bmatrix} D_1 & D_2 & D_3 \\ D_4 & D_5 & D_6 \\ D_7 & D_8 & D_9 \end{bmatrix},$$

where $D_1 = \mathcal{X}^\top\Phi^{(1)}T^{(1)}V^{*(1)}T\Phi\mathcal{X}$, $D_2 = \mathcal{X}^\top\Phi^{(1)}T^{(1)}(V^{*(1)}M + C^{(1)}]H\mathcal{Z}$, $D_3 = \mathcal{X}^\top\Phi^{(1)}T^{(1)}V^{*(1)}\Phi\boldsymbol{\rho}$, $D_4 = \mathcal{Z}^\top H^{(1)}(M^{(1)}V^{*(1)} + C^{(1)})T\Phi\mathcal{X}$, $D_5 = \mathcal{Z}^\top H^{(1)}[M^{(1)}V^{*(1)}M + (M^{(1)} + M)C^{(1)} + V^{\dagger(1)}]H\mathcal{Z}$, $D_6 = \mathcal{Z}^\top H^{(1)}(M^{(1)}V^{*(1)} + C^{(1)})\Phi\boldsymbol{\rho}$, $D_7 = \boldsymbol{\rho}^{\top(1)}\Phi^{(1)}V^{*(1)}T\Phi\mathcal{X}$, $D_8 = \boldsymbol{\rho}^{\top(1)}\Phi^{(1)}(V^{*(1)}M + C^{(1)})H\mathcal{Z}$, and $D_9 = \boldsymbol{\rho}^{\top(1)}\Phi^{(1)}V^{*(1)}\Phi\boldsymbol{\rho}$.

3.3 RAO'S SCORE TEST

It is noteworthy that it is possible to test whether the mean link function is logit by testing $\mathcal{H}_0 : \lambda = 1$ against $\mathcal{H}_1 : \lambda \neq 1$. This is possible because the logit link beta regression model is a particular case of our models. The score test (RAO, 1948) is particularly convenient here since all quantities in its test statistic are evaluated at the restricted maximum likelihood estimator and, as a consequence, estimation of λ is not required.

The general form of the score test statistic is $S_R = U(\tilde{\boldsymbol{\theta}})' K^{-1}(\tilde{\boldsymbol{\theta}})U(\tilde{\boldsymbol{\theta}})$, where $U(\tilde{\boldsymbol{\theta}})$ is the score vector and $K^{-1}(\tilde{\boldsymbol{\theta}})$ is Fisher's information inverse matrix, both evaluated at $\tilde{\boldsymbol{\theta}}$. When the interest lies in testing $\mathcal{H}_0 : \lambda = 1$ in the linear varying precision beta regression model with parametric link function, the score test statistic can be expressed as

$$S_R = U_\lambda(\tilde{\boldsymbol{\theta}})' \tilde{K}^{(\lambda\lambda)} U_\lambda(\tilde{\boldsymbol{\theta}}),$$

where

$$U_\lambda(\tilde{\boldsymbol{\theta}}) = \sum_{t=1}^{n} \left\{ \tilde{\phi}_t (y_t^* - \tilde{\mu}_t^*) \left[\frac{1}{2 + e^{-\tilde{\eta}_{1t}} + e^{\tilde{\eta}_{1t}}} - \frac{\log(1 + e^{\tilde{\eta}_{1t}})}{1 + e^{\tilde{\eta}_{1t}}} \right] \right\}$$

and

$$\begin{aligned} \tilde{K}^{(\lambda\lambda)} = \{ & \tilde{\boldsymbol{\rho}}^\top \tilde{\Phi}^2 \tilde{V}^* \tilde{\boldsymbol{\rho}} - [\tilde{\boldsymbol{\rho}}^\top \tilde{T}^\top (\tilde{V}^*)^\top (\tilde{\Phi}^2)^\top X \tilde{I}^{(\boldsymbol{\beta},\boldsymbol{\beta})} X^\top \tilde{\Phi}^2 \tilde{V}^* \tilde{T} \tilde{\boldsymbol{\rho}} \\ & + \tilde{\boldsymbol{\rho}}^\top \tilde{H}^\top (\tilde{M} \tilde{V}^* + \tilde{C})^\top \tilde{\Phi}^\top Z \tilde{I}^{(\boldsymbol{\gamma},\boldsymbol{\beta})} X^\top \tilde{\Phi}^2 \tilde{V}^* \tilde{T} \tilde{\boldsymbol{\rho}} \\ & + \tilde{\boldsymbol{\rho}}^\top \tilde{T}^\top (\tilde{V}^*)^\top (\tilde{\Phi}^2)^\top X \tilde{I}^{(\boldsymbol{\beta},\boldsymbol{\gamma})} Z^\top \tilde{\Phi} (\tilde{M} \tilde{V}^* + \tilde{C}) \tilde{H} \tilde{\boldsymbol{\rho}} \\ & + \tilde{\boldsymbol{\rho}}^\top \tilde{H}^\top (\tilde{M} \tilde{V}^* + \tilde{C})^\top \tilde{\Phi}^\top Z \tilde{I}^{(\boldsymbol{\gamma},\boldsymbol{\gamma})} Z^\top \tilde{\Phi} (\tilde{M} \tilde{V}^* + \tilde{C}) \tilde{H} \tilde{\boldsymbol{\rho}}] \}^{-1}. \end{aligned}$$

Under the null hypothesis and when n is large, S_R is approximately χ_1^2 distributed. For the nonlinear case, the score test statistic remains the same, only replacing X and Z with $\mathcal{X}$ and $\mathcal{Z}$, respectively.

4 NUMERICAL RESULTS

In this chapter we shall report the results of several Monte Carlo simulations that were performed to evaluate the finite sample performances of the likelihood ratio test (w) and of its two corrected versions (w^* and w^{**}) along with results for the score test in linear beta regressions with parametric mean link function. Since the link function parameter and the precision parameter are constrained to be positive, log-likelihood maximization was carried out using the sequential quadratic programming (SQP) nonlinear optimization algorithm with first analytical derivatives (NOCEDAL; WRIGHT, 2006, Chapter 18). It allows for the specification of linear and non-linear restrictions through sequential quadratic programming. The algorithm uses a second-order correction step in which the quasi-Newton formula updates the Hessian approximation in each iteration. In all numerical experiments, we specified the lower bound of λ as 0.001. The same lower bound was used for ϕ under constant precision. When working with corrections to the signed likelihood ratio test statistic and in order to avoid numerical instability, Severini (2000, p. 241-244) recommends that the modified test statistics be set equal to the standard test statistic when the latter is close to zero. We shall proceed similarly: w^* and w^{**} are set equal to w when the latter is ≤ 0.1. The values of all mean covariates were randomly generated from the standard uniform distribution. Under varying precision, the precision covariates are set equal to the corresponding mean covariates. The significance levels (α) are 10%, 5%, and 1%. All results are based on 10,000 Monte Carlo replications and all simulations were performed using the Ox matrix programming language (DOORNIK, 2009); see <https://www.doornik.com>.

4.1 FIXED PRECISION BETA REGRESSION WITH PARAMETRIC MEAN LINK FUNCTION

At the outset, we consider the fixed dispersion precision beta regression model with parametric link function given by

$$g(\mu_t, \lambda) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4},$$

$t = 1, \ldots, n$, where $g(\mu_t, \lambda)$ is the AO link function. We consider three different null hypotheses, namely: (i) $\mathcal{H}_0 : \lambda = 1$ (logit link), (ii) $\mathcal{H}_0 : \beta_4 = 0$, and (iii) $\mathcal{H}_0 : \lambda = 1, \beta_4 = 0$. In the first case, the parameter values are $\beta_1 = -1.5$, $\beta_2 = 1.5$, $\beta_3 = 4.0$, $\beta_4 = -4.0$, and $\lambda = 1$. For the second case, we have $\beta_1 = -1.5$, $\beta_2 = 1.5$, $\beta_3 = 4.0$, $\beta_4 = 0$, and $\lambda = 0.5$. In the third and final case, we set $\beta_1 = -1.5$, $\beta_2 = 1.5$, $\beta_3 = 4.0$, $\beta_4 = 0$, and $\lambda = 1$. The precision parameter (ϕ) equals 10 and

30, and $n \in \{20, 40, 60, 80\}$.

The tests null rejection rates (%) are presented in Table 1. First, notice that the likelihood ratio test is considerably liberal (oversized) when the sample is small. For example, when testing $\mathcal{H}_0 : \beta_4 = 0$, for $n = 20$, $\phi = 30$, and $\alpha = 10\%$, its null rejection rate is 16.1%. Second, the two corrected tests (w^* and w^{**}) display better control of the type I error frequency than the standard likelihood ratio test. Under the same conditions, their null rejection rates are 10.1% (w^*) and 10.6% (w^{**}). Overall, the two variants of the likelihood ratio test perform similarly, w^* being slightly conservative.

We also performed simulations to evaluate the tests powers, i.e., their ability to detect that the null hypothesis is false. We test $\mathcal{H}_0 : \lambda = 1$. Data generation is carried out using link parameter values that are different from one and range from 0.2 to 2.0 in increments of 0.2. The sample size is $n = 40$ and $\phi = 30$. The tests non-null rejection rates (%) can be found in Table 2. As expected, the tests become more powerful as the true value of λ moves away from one. Interestingly, the tests are more powerful when $\lambda < 1$ than when $\lambda > 1$. For instance, the tests non-null rejection rates at the 10% significance level for $\lambda = 0.4$ ($\lambda = 1.6$) range from approximately 74% to nearly 77% (from approximately 24% to 30%). We also note that the likelihood ratio test tends to be slightly more powerful than its two modified version. This advantage stems from the fact that w is oversized.

Figure 2 contains quantile-quantile (QQ) plots of the three test statistics for testing $\mathcal{H}_0 : \lambda = 1$ and $\mathcal{H}_0 : \beta_4 = 0$. The sample sizes are $n \in \{20, 40, 60\}$. We plot the exact quantiles of each test statistic against the asymptotic quantiles. The 45 degree line indicates perfect agreement between exact and asymptotic quantiles. QQ plots allow us to investigate the quality of the χ^2 asymptotic approximation at the entire distribution range, and not only at some selected quantiles (e.g., 0.90, 0.95, and 0.99). It is clear from the plots that the null distributions of w^* and w^{**} are much better approximated by the limiting null distribution than that of w. We also note that the null distribution of w^* is slightly better approximated by its limiting counterpart than that of w^{**}.

4.2 VARYING PRECISION BETA REGRESSION WITH PARAMETRIC MEAN LINK FUNCTION

We shall now consider varying precision. In particular, we shall use the following model as the true data generating process:

$$g(\mu_t, \lambda) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4}$$

Table 1 – Null rejection rates (%), fixed precision.

			10%			5%			1%		
ϕ	$\mathcal{H}_0$	n	w	w^*	w^{**}	w	w^*	w^{**}	w	w^*	w^{**}
10	$\lambda = 1$	20	15.5	9.5	9.9	9.0	4.3	4.6	2.4	0.7	0.8
		40	11.8	9.6	9.7	6.0	4.4	4.5	1.3	0.8	0.8
		60	11.2	9.9	10.0	5.8	5.0	5.0	1.2	0.8	0.8
		80	11.2	10.5	10.6	5.8	5.5	5.5	1.3	1.2	1.2
	$\beta_4 = 0$	20	15.8	10.1	10.6	9.2	5.4	6.0	2.9	1.0	1.3
		40	13.0	10.0	10.2	7.0	4.9	5.1	1.6	1.0	1.0
		60	11.3	9.4	9.7	6.2	5.1	5.3	1.3	0.9	1.0
		80	11.1	10.0	10.0	5.7	4.8	4.8	1.2	1.0	1.0
	$\lambda = 1, \beta_4 = 0$	20	15.9	12.1	13.5	8.9	6.2	7.6	2.1	0.8	1.6
		40	14.0	10.8	11.0	7.5	5.3	5.4	1.6	1.0	1.1
		60	12.7	10.4	10.6	6.6	5.1	5.2	1.3	0.9	0.9
		80	11.1	10.0	10.0	5.9	5.0	5.0	1.4	1.1	1.1
30	$\lambda = 1$	20	16.6	10.5	11.0	10.0	5.2	5.6	3.0	1.0	1.1
		40	12.2	9.6	9.7	6.8	4.9	5.0	1.9	1.1	1.1
		60	12.4	10.4	10.5	6.4	5.2	5.2	1.5	1.0	1.0
		80	11.4	10.1	10.1	6.0	5.2	5.2	1.3	1.0	1.0
	$\beta_4 = 0$	20	16.1	10.1	10.6	9.2	5.2	5.6	2.9	1.1	1.2
		40	12.2	9.8	9.9	6.6	5.1	5.1	1.5	1.0	1.0
		60	11.3	9.5	9.6	6.0	5.1	5.1	1.4	1.0	1.0
		80	11.4	10.1	10.2	6.0	5.0	5.0	1.3	1.0	1.0
	$\lambda = 1, \beta_4 = 0$	20	18.2	10.6	11.4	10.7	5.5	6.0	2.9	0.9	1.2
		40	13.8	10.2	10.4	7.4	5.3	5.4	1.8	1.0	1.1
		60	12.2	9.9	10.0	6.3	4.8	4.8	1.5	1.0	1.0
		80	11.6	10.0	10.1	6.0	4.9	4.9	1.4	1.0	1.0

Source: Author (2020)

$$\log(\phi_t) = \gamma_1 + \gamma_2 z_{t2} + \gamma_3 z_{t3} + \gamma_4 z_{t4}.$$

The null hypotheses under test are the same as in the previous simulations and also three additional ones, namely: (i) $\mathcal{H}_0 : \gamma_2 = \gamma_3 = \gamma_4 = 0$ (fixed precision), (ii) $\mathcal{H}_0 : \lambda = 1$, $\gamma_2 = \gamma_3 = \gamma_4 = 0$, and (iii) $\mathcal{H}_0 : \lambda = 1$, $\beta_4 = \gamma_2 = \gamma_3 = \gamma_4 = 0$. We shall refer to (i), (ii), and (iii) as Scenarios 1, 2, and 3, respectively. Notice that in Scenario 1 we test the null hypothesis of fixed precision. The values of the parameters in the mean submodel are $\beta_1 = -1.5$, $\beta_2 = 1.5$, $\beta_3 = 4.0$, and $\beta_4 = -4.0$. Additionally, $\gamma_2 = \gamma_3 = \gamma_4 = 0$ and we consider two values for γ_1, namely: log(10) and log(30). The sample sizes are $n \in \{30, 50, 70, 90\}$.

The null rejection rates (%) are presented in Tables 3 (Scenarios 1, 2, and 3) and 4 (null hypotheses as in the previous set of simulations). Again, the likelihood ratio test is considerably

Table 2 – Non-null rejection rates (%), fixed precision (ϕ = 30) and n = 40.

	10%			5%		
λ	w	w^*	w^{**}	w	w^*	w^{**}
0.2	98.7	98.5	98.5	97.3	96.8	96.8
0.4	76.6	74.3	74.3	66.8	63.5	63.5
0.6	37.8	35.0	35.1	27.5	24.5	24.6
0.8	17.3	15.0	15.0	10.3	8.4	8.5
1.2	15.1	11.6	11.8	8.8	6.2	6.3
1.4	24.0	18.9	19.2	15.1	11.3	11.4
1.6	30.1	24.3	24.5	20.4	15.2	15.5
1.8	36.8	30.3	30.6	25.8	20.2	20.5
2.0	43.8	36.7	37.2	31.9	25.3	25.6

Source: Author (2020)

liberal and is clearly outperformed by the two modified tests. Consider, for instance, the test of the null hypothesis that the link function is logit ($\mathcal{H}_0 : \lambda = 1$), $n = 50$, $\gamma_1 = \log(10)$, and $\alpha = 5\%$ (Table 4). The estimated sizes of w, w^*, and w^{**} are 9.1%, 5.6%, and 5.9%, respectively. The likelihood ratio test rejection rate exceeds the significance level by over 80%. The test thus rejects the true logit link function with excessive frequency. Consider now Scenario 2, $n = 30$, $\alpha = 10\%$, and the largest value of γ_1. The null rejection rates of w, w^*, and w^{**} are 25.6%, 9.0%, and 10.5%, respectively. The null rejection rate of the uncorrected test is over 2.5 times larger than the significance level. This is a very large size distortion.

We also performed power simulations under varying precision, i.e., simulations in which the true data generating process differs from that specified in the null hypothesis under evaluation. The tests non-null rejection rates (%) are displayed in Tables 5 and 6. In both cases, the sample size is $n = 70$, $\gamma_1 = \log(30)$, and we test $\mathcal{H}_0 : \lambda = 1$, i.e., we test that the mean link function is logit. At the outset, we consider a grid of values for the true value of λ similar to that used in the previous set of power simulations (Table 2, fixed precision). Again, the tests are more powerful for values of λ that are smaller than one relative to values of λ in excess of one. The unmodified test again tends to be slightly more powerful than the corrected tests. This is due to its liberal (oversized) behavior. In the second set of power simulations, we perform data generation using four well known link functions, namely: loglog, cloglog, Cauchy, and probit. The estimated powers (%) are presented in Table 6. The tests are very powerful against the loglog link function, their non-null rejection rates at the 10% significance level exceeding 87%. The tests are also quite powerful when the true link function is cloglog, their nonnull rejection rates at the 10% significance level being above 79%. There is also good power against the Cauchy link; here, their estimated powers are approximately 73% at the 10% significance level. The powers are, however, considerably lower when the true mean link function is probit, ranging from 30.6% to 32.4% when $\alpha = 10\%$. This is understandable since the logit and probit link functions are similar and a much larger sample size would be required for the tests to be able to reliably distinguish between them.

Figure 3 contains QQ plots for the test statistics when we test that (i) the mean link function is logit (one restriction) and (ii) the mean link function is logit and $\beta_4 = 0$ (two restrictions) in the varying precision beta regression model. The sample sizes are $n \in \{30, 50, 70\}$. It is clear from these plots that the null distribution of likelihood ratio test is poorly approximated by the limiting χ^2 distribution in all cases. The exact quantiles of w are considerably larger than the corresponding χ^2 quantiles and that renders the test to be considerably oversized. The χ^2 approximation holds much better when applied to the null distributions of w^* and w^{**}, especially the former. The agreement between exact and asymptotic quantiles is very good for both modified test statistics, especially when $n > 50$.

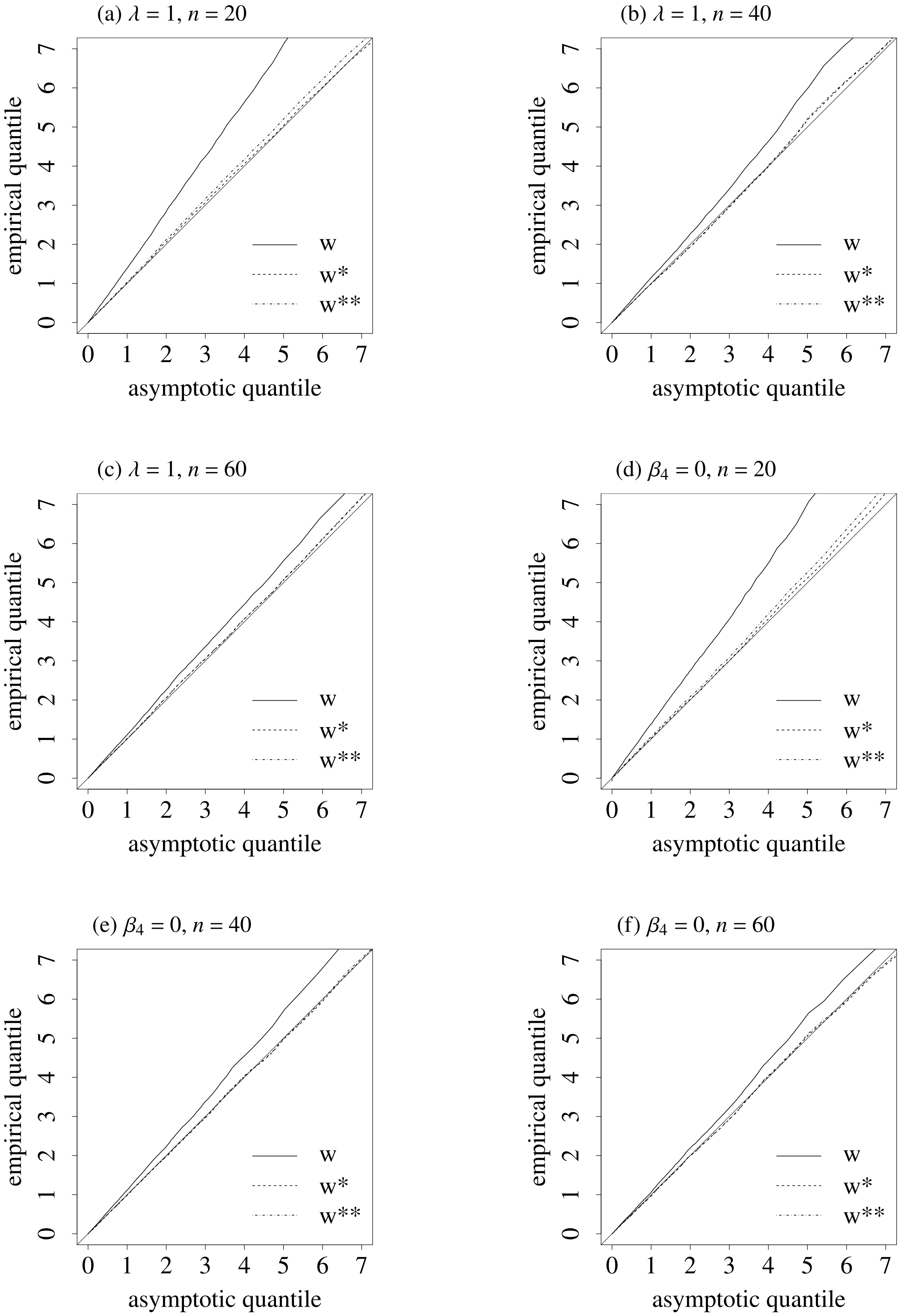
Finally, we consider inference based on the score test. The interest lies in testing $\mathcal{H}_0 : \lambda = 1$ (logit link). As noted earlier, the score test is particularly appealing for testing the logit link since it does not require estimation of λ. Table 7 displays the null rejection rates (%) of the score test. Note that for fixed (varying) precision, with $n = 20$ ($n = 30$), and $\phi = 30$ ($\gamma_1 = \log(30)$) the test is liberal, its estimated size being 14.6% (13.3%) at $\alpha = 10\%$. When $n \geq 40$, the null rejection rates are close to the nominal levels. For example, when $n = 50$, $\alpha = 5\%$, and $\gamma_1 = \log(30)$ (varying precision), the size distortion is 0.4%. Figure 4 contains QQ plots for the score test statistic (S_R) and the two modified likelihood ratio test statistics. It is noteworthy that the null distributions of w^* and w^{**} are better approximated by the limiting χ^2_1 distribution than that of the score test statistic. Inferences based on the modified likelihood ratio tests are thus more accurate than those based on the score test.

Table 3 – Null rejection rates (%), varying precision; first set of null hypotheses.

			10%			5%			1%		
γ_1	Scenario	n	w	w^*	w^{**}	w	w^*	w^{**}	w	w^*	w^{**}
log(10)	1	30	22.9	9.4	10.6	13.9	5.0	5.8	4.8	1.2	1.6
		50	15.4	10.1	10.3	8.5	4.9	5.0	2.4	1.1	1.2
		70	12.9	9.4	9.5	6.8	4.8	4.8	1.6	1.0	1.0
		90	12.5	9.9	9.9	6.4	4.9	5.0	1.3	0.8	0.9
	2	30	23.7	9.0	10.2	14.6	4.9	5.7	5.0	1.1	1.4
		50	16.0	9.7	10.0	8.9	4.5	4.7	2.2	0.9	1.0
		70	14.3	10.3	10.5	7.9	5.2	5.3	2.0	1.0	1.0
		90	12.6	10.1	10.2	6.7	5.0	5.1	1.6	1.1	1.1
	3	30	22.6	8.7	9.9	13.7	4.7	5.5	4.4	1.5	2.0
		50	16.3	9.8	10.2	9.3	4.8	5.1	2.3	0.8	0.9
		70	14.4	10.8	11.1	8.3	5.6	5.8	1.8	1.2	1.4
		90	13.4	10.3	10.6	7.5	5.1	5.2	1.8	1.2	1.2
log(30)	1	30	24.9	9.6	11.3	16.0	4.9	5.9	5.7	1.0	1.2
		50	16.8	10.2	10.6	9.8	5.1	5.3	2.7	1.0	1.1
		70	15.1	10.4	10.6	8.3	5.2	5.2	2.0	1.1	1.1
		90	13.8	10.6	10.7	7.6	5.1	5.2	2.0	1.3	1.3
	2	30	25.6	9.0	10.5	16.4	4.7	5.7	6.0	1.0	1.3
		50	17.6	9.5	10.0	10.1	4.7	5.0	2.9	1.0	1.0
		70	14.0	9.6	9.7	7.8	4.8	4.9	1.7	0.8	0.9
		90	12.8	9.8	9.9	6.9	5.0	5.1	1.8	1.1	1.1
	3	30	25.8	8.7	9.8	16.9	4.5	5.3	5.6	1.6	1.9
		50	16.6	10.4	10.9	9.7	5.5	5.8	2.9	1.3	1.6
		70	14.1	9.8	9.9	7.7	5.1	5.2	2.0	1.0	1.0
		90	13.4	10.1	10.2	7.4	5.2	5.2	2.0	1.1	1.1

Source: Author (2020)

Figure 2 – Quantile-quantile (QQ) plots, ϕ = 30, fixed precision.



Source: Author (2020)

Table 4 – Null rejection rates (%), varying precision; second set of null hypotheses.

			10%			5%			1%		
γ_1	$\mathcal{H}_0$	n	w	w^*	w^{**}	w	w^*	w^{**}	w	w^*	w^{**}
log(10)	$\lambda = 1$	30	21.4	11.5	13.2	13.5	5.8	6.9	4.4	1.4	1.6
		50	15.8	11.2	11.5	9.1	5.6	5.9	2.3	1.0	1.1
		70	13.8	10.4	10.7	7.9	5.3	5.4	2.2	1.0	1.1
		90	13.0	10.6	10.7	6.9	5.0	5.1	1.8	1.2	1.2
	$\beta_4 = 0$	30	21.9	14.1	16.2	14.3	8.1	9.7	5.2	2.1	3.1
		50	14.8	10.2	10.6	8.7	5.6	5.8	2.8	1.3	1.4
		70	13.3	10.5	10.7	7.2	5.2	5.3	2.1	1.1	1.2
		90	13.0	10.2	10.4	6.7	5.0	5.0	1.8	1.2	1.2
	$\lambda = 1, \beta_4 = 0$	30	22.5	12.9	14.5	13.8	6.8	8.3	4.4	1.5	2.3
		50	16.9	11.4	11.9	9.8	5.9	6.2	2.9	1.1	1.3
		70	14.8	11.1	11.4	8.6	5.6	5.7	2.1	1.0	1.1
		90	13.9	10.9	11.0	7.5	5.5	5.6	2.0	1.4	1.4
log(30)	$\lambda = 1$	30	20.7	12.6	13.8	13.2	7.0	7.7	4.6	1.8	2.1
		50	15.7	11.3	11.5	9.2	5.8	6.0	2.3	1.1	1.2
		70	13.1	9.7	9.9	7.0	5.0	5.1	2.0	1.1	1.1
		90	13.0	10.6	10.7	7.1	5.3	5.4	1.5	1.0	1.0
	$\beta_4 = 0$	30	20.8	13.7	15.0	13.6	7.6	8.9	4.8	2.0	2.6
		50	14.6	10.6	10.9	8.4	5.4	5.6	2.3	1.2	1.3
		70	13.1	10.3	10.4	7.5	5.4	5.5	2.1	1.2	1.2
		90	12.0	10.0	10.0	6.3	4.9	5.0	1.5	1.0	1.0
	$\lambda = 1, \beta_4 = 0$	30	25.3	13.5	15.0	16.4	7.4	8.5	6.0	1.9	2.4
		50	16.2	10.8	11.1	9.2	5.4	5.5	2.3	1.1	1.1
		70	14.1	10.0	10.3	7.7	4.9	5.0	1.9	0.9	0.9
		90	12.9	9.7	9.8	7.0	4.9	5.0	1.8	1.0	1.0

Source: Author (2020)

Table 5 – Non-null rejection rates (%), varying precision, n = 70, and γ_1 = log(30); grid of values for λ.

	10%			5%		
λ	w	w^*	w^{**}	w	w^*	w^{**}
0.2	99.3	98.9	98.9	98.4	97.4	97.4
0.4	87.2	82.6	83.0	79.0	72.8	73.2
0.6	50.8	44.6	45.0	39.6	32.3	32.7
0.8	21.7	17.2	17.4	13.4	9.7	9.9
1.2	18.6	15.2	15.4	11.2	8.5	8.6
1.4	28.8	24.7	24.8	19.2	15.4	15.6
1.6	40.2	35.9	36.1	29.1	24.5	24.7
1.8	51.0	46.6	46.9	39.5	33.9	34.1
2.0	59.3	55.0	55.3	48.0	42.4	42.6

Source: Author (2020)

Table 6 – Non-null rejection rates (%), varying precision, n = 70, and γ_1 = log(30); alternative well known link functions.

	10%			5%		
True link	w	w^*	w^{**}	w	w^*	w^{**}
loglog	91.8	87.6	88.1	86.1	78.6	79.2
cloglog	84.4	79.4	79.6	73.0	65.0	65.3
Cauchy	74.0	73.0	73.1	59.9	57.2	57.2
probit	32.4	30.6	31.0	23.6	21.8	22.2

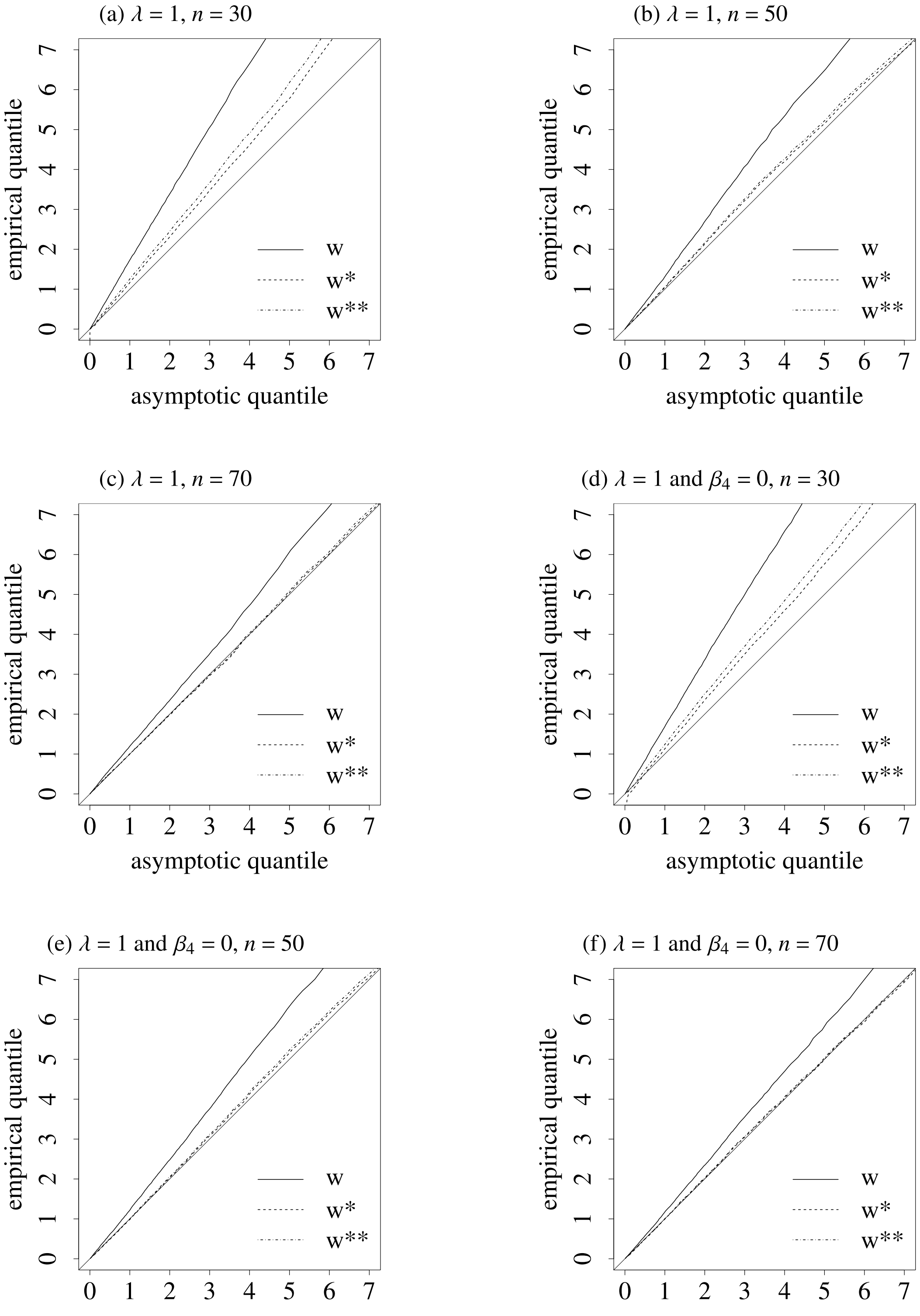
Source: Author (2020)

Table 7 – Null rejection rates (%), fixed and varying precision; score test.

Fixed precision					Varying precision				
		S_R					S_R		
ϕ	n	10%	5%	1%	γ_1	n	10%	5%	1%
10	20	13.2	7.0	1.1	log(10)	30	12.5	6.3	1.2
	40	10.9	5.4	1.2		50	10.7	5.4	1.2
	60	10.5	5.5	0.9		70	10.9	5.8	1.2
	80	10.7	5.6	1.2		90	10.7	5.2	1.1
30	20	14.6	8.0	1.5	log(30)	30	13.3	6.6	1.2
	40	11.3	6.0	1.2		50	11.7	6.1	1.0
	60	11.6	5.8	1.2		70	10.8	5.4	1.0
	80	11.0	5.6	1.1		90	11.4	5.7	1.0

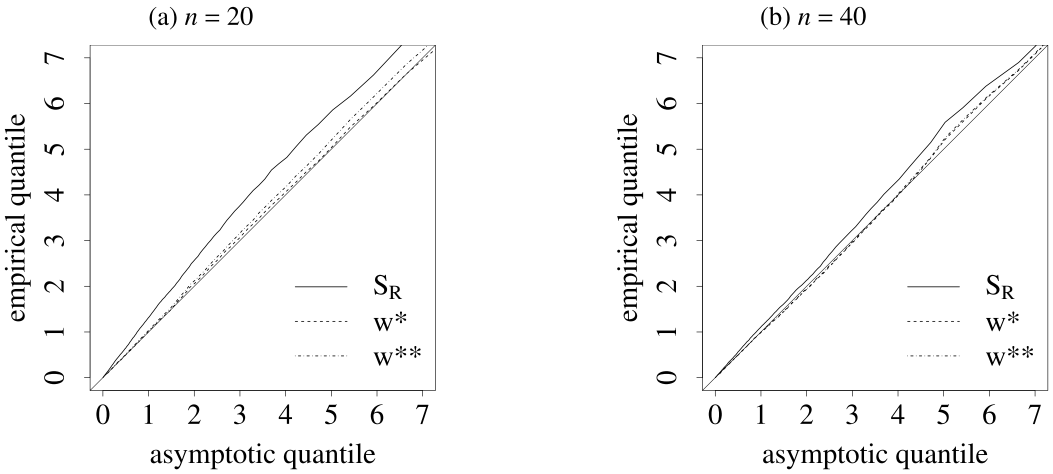
Source: Author (2020)

Figure 3 – Quantile-quantile (QQ) plots, γ_1 = log(30), varying precision.



Source: Author (2020)

Figure 4 – Quantile-quantile (QQ) plots, fixed precision (ϕ = 30), score test ($\mathcal{H}_0 : \lambda = 1$).



Source: Author (2020)

5 EMPIRICAL APPLICATIONS

In this chapter we shall present and discuss two empirical applications. The first application showcases the usefulness of the modified likelihood ratio test statistics derived in Chapter 3 whereas the second application illustrates the usefulness of the parametric mean link function when used in the beta regression model, as presented in Chapter 2. In both cases, a superior fit is achieved when the mean parametric link function is adopted. We note that, since our model is indexed by a precision (not dispersion) parameter, the results we report are directly comparable to the corresponding results reported in the literature, which use the same scale parameterization.

5.1 FIRST APPLICATION

In the first empirical application, we shall use the gasoline yield data from Prater (1956). The response variable (y) is the proportion of crude oil converted to gasoline after distillation and fractionation and $n = 32$. Ferrari and Cribari-Neto (2004) modeled the data using two covariates and the logit link function in a fixed precision beta regression. The covariate $batch$ is a dummy variable indicating ten factors of conditions in the experiments ($x_{t2},\ldots,x_{t10}$), and the covariate $temp$ (x_{t11}) represents the temperature (degrees F) at which all gasoline has vaporized. We fitted this same beta regression model only replacing the logit link function with the parametric link function. The parameter estimates are presented in Table 8 along with the corresponding standard errors, z test statistics (given by the ratio between the point estimates and the corresponding standard errors), p-values, and lower and upper limits of 95% approximate confidence intervals. Additionally, $\hat{\phi} = 942.457$ and $\hat{\lambda} = 6.602$. The estimate of β_{11}, the coefficient associated with temperature, is nearly 2/3 larger relative to the logit link model. The positive impact of temperature on the mean proportion of crude oil converted to gasoline is thus stronger when computed from a model that incorporates a parametric link function. Also, the estimated precision ($\hat{\phi}$) becomes considerably larger: 942.457 under the data-driven link function vs. 440.278 under the logit link function.

Consider the test of $\mathcal{H}_0 : \lambda = 1$ (logit link function). The likelihood ratio test statistic w equals 23.905 (p-value < 0.001). The two corrected test statistics, w^* and w^{**}, equal 14.047 (p-value < 0.001) and 15.063 (p-value < 0.001), respectively. All three tests reject the null hypothesis at 1% of significance level. Next, we test $\mathcal{H}_0 : \lambda = 6.5$. We obtain ($p$-values in parentheses) $w = 0.005538$ (0.940), $w^* = 0.000917$ (0.975), and $w^{**} = 0.001881$ (0.965). The tests p-values are very large and the null hypothesis is not rejected at the usual significance levels.

There is thus considerable evidence against the logit link function and in favor of a link function indexed by a value of λ that is substantially larger than one.

Table 8 – Parameter estimates, standard errors (SE), z statistics, p-values, and lower (LCI) and upper (UCI) limits for approximate confidence intervals; Prater's gasoline yield data.

Parameter	Estimate	SE	z statistic	p-value	LCI	UCI
β_1	−8.800	0.696	−12.640	< 0.001	−10.164	−7.435
β_2	3.238	0.393	8.230	< 0.001	2.467	4.009
β_3	2.302	0.284	8.090	< 0.001	1.744	2.860
β_4	2.698	0.320	8.422	< 0.001	2.070	3.326
β_5	1.898	0.246	7.711	< 0.001	1.416	2.381
β_6	1.915	0.239	7.992	< 0.001	1.445	2.385
β_7	1.829	0.239	7.653	< 0.001	1.360	2.297
β_8	1.021	0.177	5.747	< 0.001	0.672	1.369
β_9	0.882	0.161	5.448	< 0.001	0.564	1.199
β_{10}	0.648	0.145	4.466	< 0.001	0.363	0.932
β_{11}	0.018	0.002	9.172	< 0.001	0.014	0.022

Source: Author (2020)

Next, we shall test the null hypothesis of fixed precision in the parametric link beta regression model. To that end, we use *temp* and *press* as precision covariates, the latter representing the vapor pressure of crude oil (lbf/in2). The extended model is fitted and we test $\mathcal{H}_0 : \gamma_2 = \gamma_3 = 0$ (fixed precision). We obtain $w = 9.144$ (0.010), $w^* = 2.236$ (0.326), and $w^{**} = 3.540$ (0.170). It is noteworthy that the usual likelihood ratio test rejects the null hypothesis at the 5% significance level whereas the two corrected versions do not reject it even at $\alpha = 0.10$. The modified tests thus indicate that precision is constant across observations when the parametric link function is used. This result shows the importance of using tests with superior behavior when the sample size is not large.

In order to determine whether the two models (logit link and parametric link) are correctly specified, we shall perform the RESET misspecification test introduced by Pereira and Cribari-Neto (2014a). The null hypothesis is that the model specification is correct and the alternative hypothesis is that it is in error. We perform the misspecification test using the standard likelihood ratio test statistic and also the two modified test statistics we derived. For the logit link function model we obtain $w = 22.408$ (< 0.001), $w^* = 13.522$ (< 0.001), and $w^{**} = 14.403$ (< 0.001). All three tests suggest that the model is incorrectly specified at the 1% significance level. When the test statistics are computed from the beta regression model with parametric link function under fixed precision, we obtain $w = 0.317$ (0.573), $w^* = 0.070$ (0.791), and $w^{**} = 0.118$ (0.731). All three p-values are large and the null hypothesis of correct model formulation is not rejected. Therefore, there is evidence that the logit link model is misspecified. In contrast, the parametric link model appears to be correctly specified. We were able to achieve correct model formulation by replacing the logit link function by a parameter-index function, thus avoiding the need to model the precision.

In Table 9 we present, for both models, the following goodness-of-fit measures: (i) R^2_{FC}: the pseudo-R^2 proposed by Ferrari and Cribari-Neto (2004), (ii) R^2_N: the generalized coefficient of determination introduced by Nagelkerke (1991), (iii) AIC: the Akaike information criterion (AKAIKE, 1974), and (iv) SIC: the Schwarz information criterion (SCHWARZ, 1978). We note that the parametric link function model fit is at least as good as that of the logit link model according to all measures. In particular, we note that the difference in AIC values is nearly 22, which can be taken as substantial evidence in favor of the beta regression model that uses the parametric mean link function (BURNHAM; ANDERSON, 2004).

Table 9 – Pseudo-R^2 (R^2_{FC} and R^2_N) and model selection criteria (AIC and SIC), gasoline yield data.

Model	AIC	SIC	R^2_{FC}	R^2_N
Parametric link	−167.50	−148.45	0.96	0.99
Logit link	−145.60	−128.00	0.96	0.97

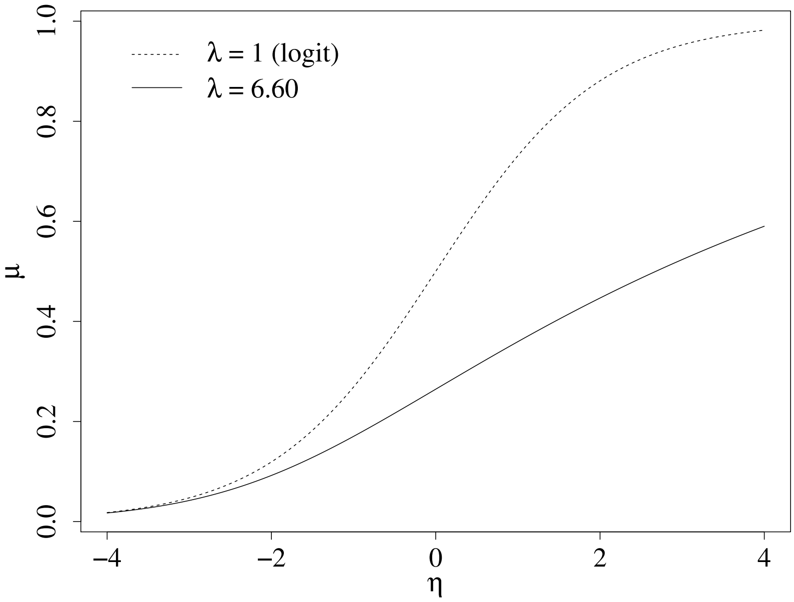
Source: Author (2020)

Finally, the link functions used in the two competing models are displayed in Figure 5. The minimum and maximum values of $\hat{\eta}_1$ in the parametric link model (logit link model) are −3.34 and 2.12 (−3.04 and 0.03), respectively, and the largest value of $\hat{\mu}$ is approximately 0.46 (0.51). It is noteworthy that the maximal response value is 0.457 (fourth observation) which is considerably closer to the corresponding predicted value obtained from the parametric link model (0.45676) than to that obtained from the logit link model (0.50792).

5.2 SECOND APPLICATION

The previous application mainly illustrated the usefulness of the two modified likelihood ratio tests. We saw that a different inference is reached regarding fixed precision when the modified tests are used. In what follows, we shall present an empirical application that illustrates the usefulness of the parametric mean link function.

Figure 5 – AO link functions, $\lambda = 1$ (logit) and $\lambda = 6.6$.



Source: Author (2020)

There is evidence that religious belief negatively correlates with intelligence; see, e.g., Lynn, Harvey and Nyborg (2009). Several authors have performed regression analyses to measure the strength of the net impact of intelligence on religious disbelief; see Zuckerman, Silberman and Hall (2013). In particular, Cribari-Neto and Souza (2013) carried out a beta regression analysis to estimate the *functional form* of such a net impact. They showed that the impact is positive, statistically significant, gains strength up to a certain level and then weakens. We shall use the same data, but using the varying precision beta regression model with parametric mean link function. The observations refer to a cross section of 124 nations. The response variable (y) is the proportion of atheists in each country and the conditioning variables are the average intelligence quotient (IQ), IQ squared (IQ^2), a dummy variable that equals 1 if the majority of the population is Muslim and 0 otherwise ($MUSL$), the logarithm of the sum of total imports and exports divided by the gross national product ($\log(OPEN)$), and the per capita income adjusted for purchasing power parity ($INCOME$).

We consider the model used by Cribari-Neto and Souza (2013) only replacing the loglog link function by the parametric link function in the mean submodel and adding a new covariate to the precision submodel ($\log(OPEN)$). Our model is

$$g(\mu_t, \lambda) = \beta_1 + \beta_2 IQ_t + \beta_3 \log(OPEN_t) + \beta_4 IQ_t^2 + \beta_5 MUSL_t + \beta_6 INCOME_t$$

$$\log(\phi_t) = \gamma_1 + \gamma_2 IQ_t + \gamma_3 \log(OPEN_t).$$

Regression coefficient estimates, standard errors, z test statistics, p-values, and lower and upper limits for 95% approximate confidence intervals can be found in Table 10. Additionally,

$\hat{\lambda} = 9.5354$. All covariates are statistically significant at the 10% significance level. We note that $MUSL$ negatively influences the mean response and that $\log(OPEN)$ and $INCOME$ have a positive impact on the mean rate of disbelief. Regarding the precision submodel, both IQ and $\log(OPEN)$ negatively affect the precision inferences. As in Cribari-Neto and Souza (2013), $\hat{\beta}_2 < 0$ and $\hat{\beta}_4 > 0$.

Table 10 – Parameter estimates, standard errors (SE), z statistics, p-values, and lower (LCI) and upper (UCI) limits for approximate confidence intervals; religious disbelief data.

	$\hat{\beta}$	SE	z statistic	p-value	LCI	UCI
Intercept	25.710	8.671	2.965	0.003	8.715	42.706
IQ	−0.918	0.235	−3.894	< 0.001	−1.381	−0.456
$\log(OPEN)$	0.717	0.205	3.488	< 0.001	0.314	1.120
IQ^2	0.006	0.001	4.091	< 0.001	0.003	0.009
$MUSL$	−0.375	0.179	−2.095	0.036	−0.726	−0.024
$INCOME$	0.029	0.015	1.820	0.068	−0.002	0.060
	$\hat{\gamma}$	SE	z statistic	p-value	LCI	UCI
Intercept	17.155	1.244	13.780	< 0.001	14.715	19.596
IQ	−0.105	0.011	−9.616	< 0.001	−0.127	−0.084
$\log(OPEN)$	−1.095	0.270	−4.044	< 0.001	−1.626	−0.564

Source: Author (2020)

We performed the RESET test of correct model specification using the three likelihood ratio test statistics. The three p-values are very large, and we conclude that the specification of our model is not in error. Goodness-of-fit measures for the parametric link function model and for the loglog link model used by Cribari-Neto and Souza (2013) are presented in Table 11. It is noteworthy that the model pseudo-R^2 values are larger for the parametric link model relative to the loglog model: for the former we obtain $R_{FC} = 0.71$ and $R_N = 0.80$ whereas for the latter we have $R_{FC} = 0.64$ and $R_N = 0.76$. Additionally, the two model selection criteria clearly favor our model: the difference in AIC between the models exceeds 18 points. According to Burnham and Anderson (2004), when the difference in the AIC values exceeds 4, one can conclude that there is considerably less support for the model with larger AIC and when it exceeds 10, it is possible to say that there is essentially no support for the model that displays the larger AIC value.

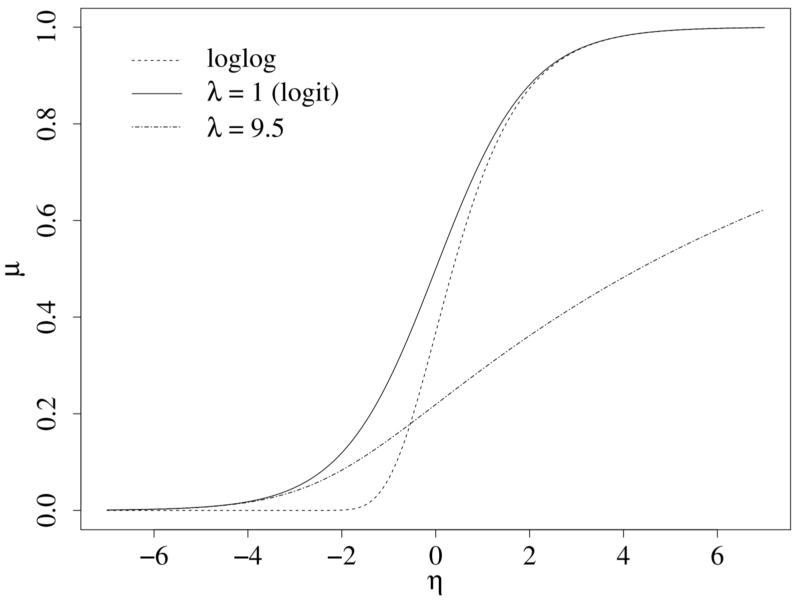
Figure 6 displays the parametric and loglog link functions; the logit link is also included for reference. The minimum and maximum values of $\hat{\eta}_1$ in the parametric link model (loglog model) are approximately −5.38 and 5.94 (−1.68 and 0.93), respectively, and the largest value of $\hat{\mu}$ is nearly 0.60 (0.67).

Table 11 – Pseudo-R^2 (R^2_{FC} and R^2_N) and model selection criteria (AIC and SIC), religious disbelief data.

Model	AIC	SIC	R^2_{FC}	R^2_N
Parametric link	−537.29	−509.09	0.71	0.80
Loglog link	−518.98	−496.42	0.64	0.76

Source: Author (2020)

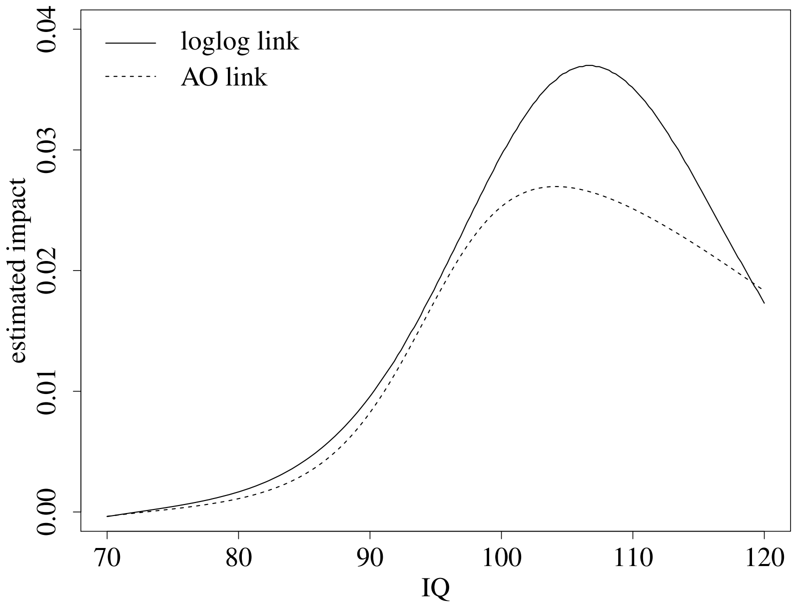
Figure 6 – AO (λ = 9.5), loglog, and logit link functions.



Source: Author (2020)

Cribari-Neto and Souza (2013) computed the impact of intelligence on the mean rate of religious disbelief by evaluating $\partial\mu_t/\partial IQ_t$ ('impact') as a function of IQ for fixed values of the remaining covariates. They showed that the impact is always positive, strengthens up to a certain level of average intelligence and then progressively looses strength. We proceeded similarly using the two models (loglog link and parametric link). Following Cribari-Neto and Souza (2013), we fix the values of all continuous regressors at the corresponding median values. The dummy variable is set at zero, i.e., we obtain estimates for non-Muslim countries. The results for Muslim countries are similar and are not presented for brevity. The impacts of average intelligence on the mean proportion of atheists computed from the two beta regression models are presented in Figure 7. In both cases, the impact is positive and bell-shaped. The maximal impact from the loglog model is 0.037 and takes place at IQ = 106. In contrast, for the parametric link beta regression model, the maximal impact equals 0.027 and corresponds to IQ = 104. The maximal impacts computed from the two models take place at approximately the same level of average intelligence, but the maximal strengths are considerably different: the maximal impact computed

Figure 7 – Impact of intelligence on atheism.



Source: Author (2020)

using the parametric link function model is over 25% weaker than that obtained from the loglog model. Interestingly, the decay after peaking is slower for the parametric link model. The use of a parametric link function thus uncovers an existing asymmetry in the rates at which the impact of intelligence on religious disbelief gains and then loses strength. Such an asymmetric behavior was not revealed in the beta regression analysis reported by (CRIBARI-NETO; SOUZA, 2013).

6 CONCLUDING REMARKS

Random variables that assume values in the standard unit interval, such as rates and proportions, are commonly modeled using the beta regression model introduced by Ferrari and Cribari-Neto (2004). We extended the varying precision version of the model by replacing its parameter-free mean link function with a parametric link. The parameter that indexes such a link function controls the degree of asymmetry and is estimated from the data. Hence, the response and covariate values determine the shape of function that connects the mean response and the linear predictor. The most commonly used link function in empirical analyses, the logit link, is a particular case of the parametric link function we use. It is thus possible to test whether it is the correct link function. We derived score functions, Fisher's information matrices and its inverses for the linear and nonlinear varying precision beta regression models with parametric mean link function.

A second contribution of the thesis relates to hypothesis testing inference. The likelihood ratio test, which is commonly used by practitioners, can be considerably size-distorted when the sample size is not large. We derived, in the context of the linear and nonlinear beta regression models with parametric mean link function, two modified versions of the likelihood ratio test statistic that yield better control of the type I error probability. We also derived a score test statistic that can be used to test whether the true mean link function is logit. Our Monte Carlo simulation evidence showed that such tests are typically much less size-distorted than the likelihood ratio test. The standard likelihood ratio test can be very liberal (oversized) in small samples. For example, in one of the Monte Carlo configurations, its null rejection at the 10% significance level exceeded 25%; the empirical sizes of our tests were 9.0% and 10.5%. This result is, we believe, indicative of the usefulness of the modified tests.

We also presented and discussed two empirical applications. In both cases, superior model fits were achieved by using a mean link function that incorporates a parameter that controls for asymmetry. In the first application, the sample size was small ($n = 32$) and the two modified tests yielded an inference regarding constant precision that was different from that obtained with the standard likelihood ratio test. In the second application, by using a parametric mean link function we were able to uncover an existing asymmetry in the impact of a regressor (average intelligence) on the mean response (mean prevalence of religious disbelievers) before and after the maximal impact. Our results revealed that the decay after peaking is slower than the strengthening prior to the peak.

BIBLIOGRAPHY

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, 1974.

ARANDA-ORDAZ, F. J. On two families of transformations to additivity for binary response data. **Biometrika**, v. 68, n. 2, p. 357–363, 1981.

BARNDORFF-NIELSEN, O. E. Inference on full or partial parameters based on the standardized signed log likelihood ratio. **Biometrika**, v. 73, n. 2, p. 307–322, 1986.

BARNDORFF-NIELSEN, O. E. Modified signed log likelihood ratio. **Biometrika**, v. 78, n. 3, p. 557–563, 1991.

BAYER, F. M.; CINTRA, R. J.; CRIBARI-NETO, F. Beta seasonal autoregressive moving average models. **Journal of Statistical Computation and Simulation**, v. 88, n. 15, p. 2961–2981, 2018.

BAYER, F. M.; CRIBARI-NETO, F. Bartlett corrections in beta regression models. **Journal of Statistical Planning and Inference**, v. 143, n. 3, p. 531–547, 2013.

BAYER, F. M.; CRIBARI-NETO, F. Bootstrap-based model selection criteria for beta regressions. **Test**, v. 24, n. 4, p. 776–795, 2015.

BAYER, F. M.; CRIBARI-NETO, F. Model selection criteria in beta regression with varying dispersion. **Communications in Statistics - Simulation and Computation**, v. 46, n. 4, p. 729–746, 2017.

BURNHAM, K. P.; ANDERSON, D. R. Multimodel inference: Understanding AIC and BIC in model selection. **Sociological Methods Research**, v. 33, n. 2, p. 261–304, 2004.

CANTERLE, D. R.; BAYER, F. M. Variable dispersion beta regressions with parametric link functions. **Statistical Papers**, v. 60, n. 5, p. 1541–1567, 2019.

COLOSIMO, E. A.; CHALITA, L. V. A. S.; DEMÉTRIO, C. G. B. Tests of proportional hazards and proportional odds models for grouped survival data. **Biometrics**, v. 56, n. 4, p. 1233–1240, 2000.

CRIBARI-NETO, F.; LUCENA, S. E. F. Non-nested hypothesis testing in the class of varying dispersion beta regressions. **Journal of Applied Statistics**, v. 42, n. 5, p. 967–985, 2015.

CRIBARI-NETO, F.; SOUZA, T. C. Religious belief and intelligence: Worldwide evidence. **Intelligence**, v. 41, n. 5, p. 482–489, 2013.

CRIBARI-NETO, F.; ZEILEIS, A. Beta regression in R. **Journal of Statistical Software**, v. 34, p. 1–24, 2010.

CZADO, C. On link selection in generalized linear models. In: **Advances in GLIM and Statistical Modelling**. New York: Springer, 1992. p. 60–65.

CZADO, C. On selecting parametric link transformation families in generalized linear models. **Journal of Statistical Planning and Inference**, v. 61, n. 1, p. 125–140, 1997.

DEHBI, H.-M.; CORTINA-BORJA, M.; GERACI, M. Aranda-Ordaz quantile regression for student performance assessment. **Journal of Applied Statistics**, v. 43, n. 1, p. 58–71, 2016.

DOORNIK, J. A. **An Object-Oriented Matrix Language Ox 6**. London: Timberlake Consultants Press, 2009.

ESPINHEIRA, P. L.; FERRARI, S. L. P.; CRIBARI-NETO, F. On beta regression residuals. **Journal of Applied Statistics**, v. 35, n. 4, p. 407–419, 2007.

ESPINHEIRA, P. L.; FERRARI, S. L. P.; CRIBARI-NETO, F. Bootstrap prediction intervals in beta regressions. **Computational Statistics**, v. 29, n. 5, p. 1263–1277, 2014.

ESPINHEIRA, P. L.; SANTOS, E. G.; CRIBARI-NETO, F. On nonlinear beta regression residuals. **Biometrical Journal**, v. 59, n. 3, p. 445–461, 2017.

FERRARI, S. L. P.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, v. 31, n. 7, p. 799–815, 2004.

FERRARI, S. L. P.; PINHEIRO, E. C. Improved likelihood inference in beta regression. **Journal of Statistical Computation and Simulation**, v. 81, n. 4, p. 431–443, 2011.

GRüN, B.; KOSMIDIS, I.; ZEILEIS, A. Extended beta regression in R: Shaken, stirred, mixed, and partitioned. **Journal of Statistical Software**, v. 48, n. 11, p. 1–25, 2012.

LAWLEY, D. N. A general method for approximating to the distribution of likelihood ratio criteria. **Biometrika**, v. 43, n. 3-4, p. 295–303, 1956.

LIMA, F. P.; CRIBARI-NETO, F. Bootstrap-based testing inference in beta regression. **Brazilian Journal of Probability and Statistics**, v. 34, n. 1, p. 18–34, 2020.

LYNN, R.; HARVEY, J.; NYBORG, H. Average intelligence predicts atheism rates across 137 nations. **Intelligence**, v. 37, n. 1, p. 11–15, 2009.

MCCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. 2. ed. London: Chapman and Hall, 1989.

NAGELKERKE, N. J. D. A note on a general definition of the coefficient of determination. **Biometrika**, v. 78, n. 3, p. 691–692, 1991.

NOCEDAL, J.; WRIGHT, S. J. **Numerical Optimization**. 2. ed. New York: Springer, 2006.

OSPINA, R.; CRIBARI-NETO, F.; VASCONCELLOS, K. L. P. Improved point and interval estimation for a beta regression model. **Computational Statistics & Data Analysis**, v. 51, n. 2, p. 960–981, 2006.

OSPINA, R.; FERRARI, S. L. P. Inflated beta distributions. **Statistical Papers**, v. 51, n. 1, p. 111–126, 2010.

OSPINA, R.; FERRARI, S. L. P. A general zero-or-one inflated beta regression models. **Computational Statistics & Data Analysis**, v. 56, n. 6, p. 1609–1623, 2012.

PAOLINO, P. Maximum likelihood estimation of models with beta-distributed dependent variables. **Political Analysis**, v. 9, n. 4, p. 325–346, 2001.

PAWITAN, Y. **In All Likelihood: Statistical Modelling and Inference Using Likelihood**. 1. ed. Oxford: Oxford University Press, 2001.

PEREIRA, T. L.; CRIBARI-NETO, F. Detecting model misspecification in inflated beta regressions. **Communications in Statistics - Simulation and Computation**, v. 43, n. 3, p. 631–656, 2014.

PEREIRA, T. L.; CRIBARI-NETO, F. Modified likelihood ratio statistics for inflated beta regressions. **Journal of Statistical Computation and Simulation**, v. 84, n. 5, p. 982–998, 2014.

PRATER, N. H. Estimate gasoline yields from crudes. **Petroleum Refiner**, v. 35, n. 5, p. 236–238, 1956.

PREGIBON, D. Goodness of link tests for generalized linear models. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, v. 29, n. 1, p. 15–24, 1980.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2018. Available at: <https://www.R-project.org/>.

RAO, C. R. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. **Mathematical Proceedings of the Cambridge Philosophical Society**, v. 44, n. 1, p. 50–57, 1948.

RAO, C. R. **Linear Statistical Inference and its Applications**. 2. ed. New York: Wiley New York, 1973.

ROCHA, A. V.; CRIBARI-NETO, F. Beta autoregressive moving average models. **Test**, v. 18, p. 529–545, 2009.

SCHER, V. T.; CRIBARI-NETO, F.; PUMI, G.; BAYER, F. M. Goodness-of-fit tests for βARMA hydrological time series modeling. **Environmetrics**, 2020. Forthcoming.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, v. 6, n. 2, p. 461–464, 1978.

SEVERINI, T. A. **Likelihood Methods in Statistics**. 1. ed. Oxford: Oxford University Press, 2000.

SIMAS, A. B.; BARRETO-SOUZA, W.; ROCHA, A. V. Improved estimators for a general class of beta regression models. **Computational Statistics & Data Analysis**, v. 54, n. 2, p. 348–366, 2010.

SKOVGAARD, I. M. Likelihood asymptotics. **Scandinavian Journal of Statistics**, v. 28, n. 1, p. 3–32, 2001.

SMITHSON, M.; VERKUILEN, J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. **Psychological Methods**, v. 11, n. 1, p. 54, 2006.

WEI, B.-C. **Exponential Family Nonlinear Models**. New York: Springer, 1998. (Lecture Notes in Statistics).

ZUCKERMAN, M.; SILBERMAN, J.; HALL, J. A. The relation between intelligence and religiosity: A meta-analysis and some proposed explanations. **Personality and Social Psychology Review**, v. 17, n. 4, p. 325–354, 2013.

APPENDIX A – COMPUTATIONAL IMPLEMENTATION

The Ox code used in the gasoline yield data empirical application is presented below.

```
/*************************************************************************************************************
*
*       PROGRAM: appPrater.ox
*
*       USAGE: Calculation of the modified test statistics for
*              the gasoline yield data from Prater (1956)
*
*       NULL HYPOTHESIS: H_0: lambda = 1
*
*       MODEL: g(mu,lambda) = X*beta
*              beta  = (beta_1...beta_p)
*              fixed precision (phi)
*
*       AUTHOR: Cristine Rauber Oliveira
*
*************************************************************************************************************/

        // header files
        #include <oxstd.oxh>
        #include <oxprob.oxh>
        #import  <maximize>
        #import  <maxsqp>

        // global variables
        static decl y;
        static decl N;
        static decl X;
        static decl Xrr;
        static decl Xrrt;
        static decl Z;
        static decl Xt;
        static decl Zt;

        // irrestricted log-likelihood function
        floglik(const vtheta, const adFunc, const avScore, const amHess)
        {
        decl r       = columns(X);
        decl beta    = vtheta[0:(r-1)];
        decl phi     = vtheta[r];
        decl lambda  = vtheta[r+1];
        decl eta1    = X*beta;
        decl mu      = 1.0 - (1.0 + lambda .* exp(eta1)) .^ (-1.0 ./ lambda);
        decl p       = mu .* phi;
        decl q       = (1.0 - mu) .* phi;

        decl ystar   = log(y ./ (1.0 - y));
        decl ydag    = log(1.0 - y);
        decl mustar  = polygamma(mu .* phi, 0) - polygamma((1.0 - mu) .* phi, 0);
        decl mudag   = polygamma((1.0 - mu) .* phi, 0) - polygamma(phi, 0);

        decl T       = diag( exp(eta1) .* (1.0 + lambda .* exp(eta1)) .^ (-(1.0 + (1.0 ./ lambda))) );
        decl H       = unit(N);
        decl P       = phi .* unit(N);
        decl M       = diag(mu);
        decl rho     = (1.0 ./ lambda) .* ((1.0 ./ (exp(-eta1) + lambda)) -
                       (log(1.0 + lambda .* exp(eta1)) ./ lambda)) .*
                       ((1.0 + lambda .* exp(eta1)) .^ (-1.0 ./ lambda));

        adFunc[0]    = double ( sumc( log(densbeta(y, p, q)) ) );

        // first order analytical derivatives of the log-likelihood function
        if(avScore)
        {
        (avScore[0])[0:(r-1)]  = Xt*P*T*(ystar - mustar);
        (avScore[0])[r]        = Zt*H*(M*(ystar - mustar) + (ydag - mudag));
        (avScore[0])[r+1]      = rho'*P*(ystar - mustar);
        }

        if( isnan(adFunc[0]) || isdotinf(adFunc[0]) )
        return 0;
```

```
else
return 1; // 1 indicates success

}

// restricted log-likelihood function
flogliknull(const vtheta, const adFunc, const avScore, const amHess)
{
decl r      = columns(X);
decl beta   = vtheta[0:(r-1)];
decl phi    = vtheta[r];
decl lambda = 1;
decl eta1   = X*beta;
decl mu     = 1.0 - (1.0 + lambda .* exp(eta1)) .^ (-1.0 ./ lambda);
decl p      = mu .* phi;
decl q      = (1.0 - mu) .* phi;

decl ystar  = log(y ./ (1.0 - y));
decl ydag   = log(1.0 - y);
decl mustar = polygamma(mu .* phi, 0) - polygamma((1.0 - mu) .* phi, 0);
decl mudag  = polygamma((1.0 - mu) .* phi, 0) - polygamma(phi, 0);

decl T      = diag( exp(eta1) .* (1.0 + lambda .* exp(eta1)) .^ (-(1.0 + (1.0 ./ lambda))) );
decl H      = unit(N);
decl P      = phi .* unit(N);
decl M      = diag(mu);

adFunc[0]   = double ( sumc( log(densbeta(y, p, q)) ) );

// first order analytical derivatives of the log-likelihood function
if(avScore)
{
(avScore[0])[0:(r-1)]  = Xt*P*T*(ystar - mustar);
(avScore[0])[r]        = Zt*H*(M*(ystar - mustar) + (ydag - mudag));
}

if( isnan(adFunc[0]) || isdotinf(adFunc[0]) )
return 0;
else
return 1; // 1 indicates success

}

// log-likelihood function of the null model
flogliknullaranda(const vtheta, const adFunc, const avScore, const amHess)
{
decl beta   = vtheta[0];
decl lambda = 1;
decl phi    = vtheta[1];
decl eta1   = Xrr*beta;
decl mu     = 1.0 - (1.0 + lambda .* exp(eta1)) .^ (-1.0 ./ lambda);
decl p      = mu .* phi;
decl q      = (1.0 - mu) .* phi;

decl ystar  = log(y ./ (1.0 - y));
decl ydag   = log(1.0 - y);
decl mustar = polygamma(mu .* phi, 0) - polygamma((1.0 - mu) .* phi, 0);
decl mudag  = polygamma((1.0 - mu) .* phi, 0) - polygamma(phi, 0);

decl T      = diag( exp(eta1) .* (1.0 + lambda .* exp(eta1)) .^ (-(1.0 + (1.0 ./ lambda))) );

adFunc[0]   = double ( sumc( log(densbeta(y, p, q)) ) );

// first order analytical derivatives of the log-likelihood function
if(avScore)
{
(avScore[0])[0] = phi*Xrrt*T*(ystar - mustar);
(avScore[0])[1] = double( sumc( mu .* (ystar - mustar) + (ydag - mudag) ) );
}

if( isnan(adFunc[0]) || isdotinf(adFunc[0]) )
return 0;
else
return 1; // 1 indicates success
```

```
}

main()
{

// variables used in the maximization of the log-likelihood function
decl dfunc0, dfunc1, dfuncr, conv0, conv1, conv2;
decl vtheta0, vtheta1, vthetar;
decl vlo1, vhi1, vlo0, vhi0, vlo2, vhi2;

// other variables used
decl ybar, yvar, ystar, ydagger;
decl r, s, k, gl, pseudoR2LR;
decl w, pvw, ws, pvws, wss, pvwss;

// variables used for the initial values
decl betaols, gamaols, phiols, varols, muols, etaols, lambdaini;

// variables used in the model
decl data, batch, temp;

oxwarning(0);
data     = loadmat("gasoline.mat");  // load the data
y        = data[][10];               // variable of interest
temp     = data[][9];                // covariate temp
batch    = data[][0:8];              // covariate batch (dummy covariates)
X        = 1~batch~temp;             // matrix 32x11
Z        = X[][0];                   // matrix 32x1 of 1's
Xt       = X';                       // X transposed
Zt       = Z';                       // Z transposed

k  = 1;                   // number of parameters of interest
r  = columns(X);          // number of parameters in the mean submodel
s  = columns(Z);          // number of parameters in the precision submodel
N  = rows(data);          // sample size
gl = N - (r + s +1)       // degrees of fredom

ystar    = log(y ./ (1.0 - y));      // transformed variable
ydagger  = log(1.0 - y);             // transformed variable

ols2c(ystar, X, &betaols);           // store the ols estimates in betaols
etaols    = X*betaols;
muols     = exp(etaols) ./ (1.0 + exp(etaols));
varols    = ((ystar - etaols)' * (ystar - etaols)) ./
            ((N - r) * ((1 ./ (muols .* (1.0 - muols))) .^ (2)));
phiols    = double( meanc((muols .* (1.0 - muols) ./ varols) - 1.0) );
lambdaini = 1;          // initial value for lambda (logit)

// initial values
vtheta1   = betaols | phiols | lambdaini;
vtheta0   = betaols | phiols;

// boundaries for the initial values
vlo1 = <-.Inf;-.Inf;-.Inf;-.Inf;-.Inf;-.Inf;-.Inf;-.Inf;-.Inf;-.Inf;-.Inf;0.001;0.001>;
vhi1 = <+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf>;
vlo0 = <-.Inf;-.Inf;-.Inf;-.Inf;-.Inf;-.Inf;-.Inf;-.Inf;-.Inf;-.Inf;-.Inf;0.001>;
vhi0 = <+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf;+.Inf>;

ybar = meanc(y);          // mean of y
yvar = varc(y);           // variance of y

println("------------------------------------------------------------------");
println("\t\t\t\t BETA REGRESSION ESTIMATION");
println("------------------------------------------------------------------");

println("\n MEAN AND VARIANCE OF Y:\n ", "%10.5f", ybar~yvar);
println("\n INITIAL VALUES FOR THE ML ESTIMATION:\n ", "%16.5f", vtheta1);
println("------------------------------------------------------------------");

// convergence checking
conv1 = MaxSQP(floglik, &vtheta1, &dfunc1, 0, 0, 0, 0, vlo1, vhi1);
conv0 = MaxSQP(flogliknull, &vtheta0, &dfunc0, 0, 0, 0, 0, vlo0, vhi0);
println("\n CONVERGENCE STATUS UNDER H1: ", MaxConvergenceMsg(conv1));
```

```
println("\n CONVERGENCE STATUS UNDER H0: ", MaxConvergenceMsg(conv0));

if(conv1 == MAX_CONV || conv1 == MAX_WEAK_CONV) && (conv0 == MAX_CONV || conv0 == MAX_WEAK_CONV))
{
decl iota        = ones(N,1);  // N-dimensional vector of ones
decl Ystar       = diag(ystar);
decl Ydagger     = diag(ydagger);
decl r           = columns(X);

// quantities under H1*******************************************************************************
decl eta1hat     = X*vtheta1[0:(r-1)];
decl phihat      = vtheta1[r];
decl lambdahat   = vtheta1[r+1];
decl muhat       = 1.0 - (1.0 + lambdahat .* exp(eta1hat)) .^ (-1.0 ./ lambdahat);

decl Hhat        = unit(N);
decl That        = diag(exp(eta1hat) .* (1.0 + lambdahat .* exp(eta1hat)) .^
                       (-(1.0 + (1.0 ./ lambdahat))));
decl Phat        = phihat .* unit(N);
decl Muhat       = diag(muhat);
decl mustarhat   = polygamma(muhat .* phihat, 0) - polygamma((1.0 - muhat) .* phihat, 0);
decl Mustarhat   = diag(mustarhat);
decl Mudaggerhat = diag(polygamma((1.0 - muhat) .* phihat, 0) - polygamma(phihat, 0));
decl vstarhat    = polygamma(muhat .* phihat, 1) + polygamma((1.0 - muhat) .* phihat, 1);
decl Vstarhat    = diag(vstarhat);
decl Vdaggerhat  = diag(polygamma((1.0 - muhat) .* phihat, 1) - polygamma(phihat, 1));
decl Chat        = diag(-polygamma((1.0 - muhat) .* phihat, 1));
decl Shat        = diag((lambdahat - lambdahat .* (1.0 + lambdahat) .* (1.0 - muhat) .^
                       (lambdahat)) ./ (((muhat - 1.0) .^ 2) .* ((1.0 - muhat) .^
                       (lambdahat) - 1.0) .^ 2));
decl Qhat        = zeros(N);
decl rhohat      = (1.0 ./ lambdahat) .* ((1.0 ./ (exp(-eta1hat) + lambdahat)) -
                   (log(1.0 + lambdahat .* exp(eta1hat)) ./ lambdahat)) .*
                   ((1.0 + lambdahat .* exp(eta1hat)) .^ (-1.0 ./ lambdahat));
decl varrhohat   = ((1.0 + lambdahat .* exp(eta1hat)) .^ (-2.0 - (1.0 ./ lambdahat)) .*
                   (-exp(eta1hat) .* (lambdahat .^ 2) .* (2.0 + exp(eta1hat) .*
                   (1.0 + 3.0 .* lambdahat)) + (1.0 + lambdahat .* exp(eta1hat)) .*
                   log(1.0 + lambdahat .* exp(eta1hat)) .* (2.0 .* lambdahat .*
                   (1.0 + exp(eta1hat) .* (1.0 + lambdahat)) - (1.0 + lambdahat .*
                   exp(eta1hat)) .* log(1.0 + lambdahat .* exp(eta1hat))))) ./ (lambdahat .^ 4);
decl what        = (exp(eta1hat) .* (1.0 + lambdahat .* exp(eta1hat)) .^
                   (-2.0 - (1.0 ./ lambdahat)) .* (-exp(eta1hat) .* lambdahat .*
                   (1.0 + lambdahat) + (1.0 + lambdahat .* exp(eta1hat)) .*
                   log(1.0 + lambdahat .* exp(eta1hat)))) ./ (lambdahat .^ 2);

// observed information******************************************************************************
decl Jbbhat = Xt*(Phat*That*Vstarhat + Shat*(That^2)*(Ystar - Mustarhat))*That*Phat*X;
decl Jbghat = -Xt*((Ystar - Mustarhat) - Phat*(Muhat*Vstarhat + Chat))*That*Hhat*Z;
decl Jblhat = Xt*((Phat^2)*Vstarhat*That*rhohat - Phat*(Ystar - Mustarhat)*what);
decl Jgbhat = Jbghat';
decl Jgghat = Zt*(Hhat*(Muhat*Vstarhat*Muhat + (Muhat + Muhat)*Chat + Vdaggerhat) +
              (Muhat*(Ystar - Mustarhat) + (Ydagger - Mudaggerhat))*(Hhat^2)*Qhat)*Hhat*Z;
decl Jglhat = -Zt*((Ystar - Mustarhat) - Phat*(Muhat*Vstarhat + Chat))*Hhat*rhohat;
decl Jlbhat = Jblhat';
decl Jlghat = Jglhat';
decl Jllhat = ((Phat^2)*Vstarhat*(rhohat .^ 2)-Phat*(Ystar - Mustarhat)*varrhohat)'*iota;

decl Jhat    = (Jbbhat~Jbghat~Jblhat) | (Jgbhat~Jgghat~Jglhat) | (Jlbhat~Jlghat~Jllhat);
decl invJhat = invert(Jhat);  // inverse of Jhat

// Fisher's information******************************************************************************
decl Kbbhat = Xt*Phat*That*Vstarhat*That*Phat*X;
decl Kbghat = Xt*Phat*(Muhat*Vstarhat + Chat)*Hhat*That*Z;
decl Kblhat = Xt*Phat*Vstarhat*Phat*That*rhohat;
decl Kgbhat = Kbghat';
decl Kgghat = Zt*Hhat*(Muhat*Vstarhat*Muhat + (Muhat+Muhat)*Chat + Vdaggerhat)*Hhat*Z;
decl Kglhat = Zt*Phat*(Muhat*Vstarhat + Chat)*Hhat*rhohat;
decl Klbhat = Kblhat';
decl Klghat = Kglhat';
decl Kllhat = rhohat'*(Phat^2)*Vstarhat*rhohat;

decl Khat    = (Kbbhat~Kbghat~Kblhat) | (Kgbhat~Kgghat~Kglhat) | (Klbhat~Klghat~Kllhat);
decl invKhat = invert(Khat); // inverse of Khat
```

```
// quantities under H0****************************************************************************
decl eta1til     = X*vtheta0[0:(r-1)];
decl phitil      = vtheta0[r];
decl lambdatil   = 1;
decl mutil       = 1.0 - (1.0 + lambdatil .* exp(eta1til)) .^ (-1.0 ./ lambdatil);

decl Htil        = unit(N);
decl Ttil        = diag(exp(eta1til) .* (1.0 + lambdatil .* exp(eta1til)) .^
                        (-(1.0 + (1.0 ./ lambdatil))));
decl Ptil        = phitil .* unit(N);
decl Mutil       = diag(mutil);
decl mustartil   = polygamma(mutil .* phitil, 0) - polygamma((1.0 - mutil) .* phitil, 0);
decl Mustartil   = diag(mustartil);
decl mudaggertil = polygamma((1.0 - mutil) .* phitil, 0) - polygamma(phitil, 0);
decl Mudaggertil = diag(mudaggertil);
decl Vstartil    = diag(polygamma(mutil .* phitil, 1) + polygamma((1.0 - mutil) .* phitil, 1));
decl Vdaggertil  = diag(polygamma((1.0 - mutil) .* phitil, 1) - polygamma(phitil, 1));
decl Ctil        = diag(-polygamma((1.0 - mutil) .* phitil, 1));
decl Stil        = diag((lambdatil - lambdatil .* (1.0 + lambdatil) .* (1.0 - mutil) .^
                        (lambdatil)) ./ (((mutil - 1.0) .^ 2) .* ((1.0 - mutil) .^
                        (lambdatil) - 1.0) .^ 2));
decl Qtil        = zeros(N);
decl rhotil      = (1.0 ./ lambdatil) .* ((1.0 ./ (exp(-eta1til) + lambdatil)) -
                   (log(1.0 + lambdatil .* exp(eta1til)) ./ lambdatil)) .*
                   ((1.0 + lambdatil .* exp(eta1til)) .^ (-1.0 ./ lambdatil));
decl varrhotil   = ((1.0 + lambdatil .* exp(eta1til)) .^ (-2.0 - (1.0 ./ lambdatil)) .*
                   (-exp(eta1til) .* (lambdatil .^ 2) .* (2.0 + exp(eta1til) .*
                   (1.0 + 3.0 .* lambdatil)) + (1.0 + lambdatil .* exp(eta1til)) .*
                   log(1.0 + lambdatil .* exp(eta1til)) .* (2.0 .* lambdatil .*
                   (1.0 + exp(eta1til) .* (1.0 + lambdatil)) - (1.0 + lambdatil .*
                   exp(eta1til)) .* log(1.0 + lambdatil .* exp(eta1til))))) ./ (lambdatil .^ 4);
decl wtil        = (exp(eta1til) .* (1.0 + lambdatil .* exp(eta1til)) .^
                   (-2.0 - (1.0 ./ lambdatil)) .* (-exp(eta1til) .* lambdatil .*
                   (1.0 + lambdatil) + (1.0 + lambdatil .* exp(eta1til)) .*
                   log(1.0 + lambdatil .* exp(eta1til)))) ./ (lambdatil .^ 2);

// observed information***************************************************************************
decl Jbbtil = Xt*(Ptil*Ttil*Vstartil + Ttil*Stil*Ttil*(Ystar - Mustartil))*Ttil*Ptil*X;
decl Jbgtil = -Xt*((Ystar - Mustartil) - Ptil*(Mutil*Vstartil + Ctil))*Ttil*Htil*Z;
decl Jbltil = Xt*(Ptil*Vstartil*Ptil*Ttil*rhotil - Ptil*(Ystar - Mustartil)*wtil);
decl Jgbtil = Jbgtil';
decl Jggtil = Zt*(Htil*(Mutil*Vstartil*Mutil + (Mutil + Mutil)*Ctil + Vdaggertil) +
              (Mutil*(Ystar - Mustartil) + (Ydagger - Mudaggertil))*Htil*Qtil*Htil)*Htil*Z;
decl Jgltil = -Zt*((Ystar - Mustartil) - Ptil*(Mutil*Vstartil + Ctil))*Htil*rhotil;
decl Jlbtil = Jbltil';
decl Jlgtil = Jgltil';
decl Jlltil = ((Ptil .^ 2)*Vstartil*(rhotil .^ 2) - Ptil*(Ystar - Mustartil)*varrhotil)'*iota;

decl Jtil    = (Jbbtil~Jbgtil~Jbltil) | (Jgbtil~Jggtil~Jgltil) | (Jlbtil~Jlgtil~Jlltil);
decl invJtil = invert(Jtil); // inverse  of Jtil

// Fisher's information***************************************************************************
decl Kbbtil = Xt*Ptil*Ttil*Vstartil*Ttil*Ptil*X;
decl Kbgtil = Xt*Ptil*(Mutil*Vstartil + Ctil)*Ttil*Htil*Z;
decl Kbltil = Xt*Ptil*Vstartil*Ptil*Ttil*rhotil;
decl Kgbtil = Kbgtil';
decl Kggtil = Zt*Htil*(Mutil*Vstartil*Mutil + (Mutil + Mutil)*Ctil + Vdaggertil)*Htil*Z;
decl Kgltil = Zt*Ptil*(Mutil*Vstartil + Ctil)*Htil*rhotil;
decl Klbtil = Kbltil';
decl Klgtil = Kgltil';
decl Klltil = rhotil'*(Ptil .^ 2)*Vstartil*rhotil;

decl Ktil    = (Kbbtil~Kbgtil~Kbltil) | (Kgbtil~Kggtil~Kgltil) | (Klbtil~Klgtil~Klltil);
decl invKtil = invert(Ktil);  // inverse of Ktil

// score function under H0************************************************************************
decl escorebetatil    = Xt*Ptil*Ttil*(ystar - mustartil);
decl escoregamatil    = Zt*Htil*(Mutil*(ystar - mustartil) + (ydagger - mudaggertil));
decl escorelambdatil  = rhotil'*Ptil*(ystar - mustartil);

decl escoretil        = escorebetatil | escoregamatil | escorelambdatil;

// qbar*******************************************************************************************
decl qbeta   = Xt*Phat*That*(Vstarhat*(Phat*Muhat - Ptil*Mutil) + (Phat - Ptil)*Chat)*iota;
```

```
decl qgama   = Zt*Hhat*((Muhat*Vstarhat + Chat)*(Phat*Muhat - Ptil*Mutil) +
                        (Muhat*Chat + Vdaggerhat)*(Phat - Ptil))*iota;
decl qlambda = rhohat'*Phat*(Vstarhat*(Phat*Muhat - Ptil*Mutil) + Chat*(Phat - Ptil))*iota;

decl qbar    = qbeta | qgama | qlambda;

// upsilonbar****************************************************************************************
decl upsbb = Xt*Phat*That*Vstarhat*Ttil*Ptil*X;
decl upsbg = Xt*Phat*That*(Vstarhat*Mutil + Chat)*Htil*Z;
decl upsbl = Xt*Phat*That*Vstarhat*Ptil*rhotil;
decl upsgb = Zt*Hhat*(Muhat*Vstarhat + Chat)*Ttil*Ptil*X;
decl upsgg = Zt*Hhat*(Muhat*Vstarhat*Mutil + (Muhat + Mutil)*Chat + Vdaggerhat)*Htil*Z;
decl upsgl = Zt*Hhat*(Muhat*Vstarhat + Chat)*Ptil*rhotil;
decl upslb = rhohat'*Phat*Vstarhat*Ttil*Ptil*X;
decl upslg = rhohat'*Phat*(Vstarhat*Mutil + Chat)*Htil*Z;
decl upsll = rhohat'*Phat*Vstarhat*Ptil*rhotil;

decl upsbar     = (upsbb~upsbg~upsbl) | (upsgb~upsgg~upsgl) | (upslb~upslg~upsll);
decl invupsbar  = invert(upsbar); // inverse of upsbar

decl nuiJtil    = Jtil[0:(r+s-1)][0:(r+s-1)];
decl nuisance   = Ktil*invupsbar*Jhat*invKhat*upsbar;
decl nuisance2  = nuisance[0:(r+s-1)][0:(r+s-1)];

// Likelihood ratio test statistics******************************************************************

w   = 2*(dfunc1 - dfunc0);          // likelihood ratio test statistic
pvw = 1.0 - probchi(w,1);           // p-value of w

decl epson  = fabs(((fabs(determinant(Ktil))*fabs(determinant(Khat))*
                    fabs(determinant(nuiJtil)))^(0.5))/(fabs(determinant(upsbar))*
                    fabs(determinant(nuisance2))^(0.5))*
                   (fabs(escoretil'*invupsbar*Khat*invJhat*upsbar*invKtil*escoretil)^(1/2))/
                    fabs((w)^((1/2)-1.0)*escoretil'*invupsbar*qbar));

ws    = w - 2*log(epson);           // Skovgaard's modified likelihood ratio test statistic w*
pvws  = 1.0 - probchi(ws,1);        // p-value of w*
wss   = w*(1.0 - log(epson)/w)^2;   // Skovgaard's modified likelihood ratio test statistic w**
pvwss = 1.0 - probchi(wss,1);       // p-value of w**

// pseudoR2 based on likelihood**********************************************************************
Xrr           = ones(N, 1);
Xrrt          = Xrr';
decl ystarbar = meanc(ystar);
decl lambdar  = 1;
decl muhatr   = 1.0 - (1.0 + lambdar .* exp(ystarbar)) .^ (-1.0 ./ lambdar);

// initial values (constant mean and precision, fixed lambda)****************************************
vthetar       = ystarbar | ((1.0/(varc(ystar)*muhatr*(1.0 - muhatr))));
vlo2          = <-.Inf;-.Inf>;
vhi2          = <+.Inf;+.Inf>;

// convergence checking
conv2 = MaxSQP(flogliknullaranda, &vthetar, &dfuncr, 0, 0, 0, 0, vlo2, vhi2);

if(conv2 == MAX_CONV || conv2 == MAX_WEAK_CONV)
{
pseudoR2LR  = 1.0 - (exp(dfuncr)/exp(dfunc1))^(2/N);
}

// measures of quality of the fitted model***********************************************************
decl pseudoR2  = (correlation(eta1hat~ystar)[0][1])^2;
decl AIC       = -2*dfunc1 + 2*(r + s + 1);
decl BIC       = -2*dfunc1 + (r + s + 1)*log(N);

// printing results**********************************************************************************
println("\n PARAMETER ESTIMATES AND ASYMPTOTIC STANDARD ERRORS: ");
decl stderrors = sqrt(diagonal(invKhat))';             // standard erros
decl zstats    = vtheta1 ./ stderrors;                 // z test statistic
println("%16.5f", "%c", {"estimates", "std. errors", "z stats", "p-values"}, "%r",
                        {"intercept", "batch1", "batch2", "batch3", "batch4", "batch5",
                         "batch6", "batch7", "batch8", "batch9", "temp", "phi", "lambda"},
                          vtheta1~stderrors~zstats~2.0*(1.0-probn(fabs(zstats))));
println("\t Sample size:", N);        // sample size
```

```
println("%r", {"pseudoR2", "pseudoR2LR", "AIC", "BIC"}, pseudoR2  | pseudoR2LR | AIC | BIC);
println("\n ASYMPTOTIC COVARIANCE MATRIX OF ML ESTIMATES:");
println("%14.5f", invKhat);

println("------------------------------------------------------------------");
println("\t\t\t\t LIKELIHOOD RATIO TEST STATISTICS");
println("------------------------------------------------------------------");
println("\t\t\t NULL HYPOTHESIS: lambda = 1 (logit link):");
println("------------------------------------------------------------------");
println("%16.6f", "%c", {"\t test statistic", "p-value"}, "%r", {"w", "w*", "w**"},
                                        (w | ws | wss)~(pvw | pvws | pvwss));
}
else
{
println("\n\n ERROR: NO CONVERGENCE!\n\n");
}

println("------------------------------------------------------------------");
println("\t\t\t Program:", oxfilename(0));
println("\t\t\t OX version:", oxversion());
println("\t\t\t Maximization algorithm applied: MaxSQP");
println("\t\t\t Date:", date());
println("\t\t\t Time: ", time());
println("------------------------------------------------------------------");

}  // end of main*********************************************************************************
```